ESTUDIO COMPARATIVO ENTRE UN MODELO DE ANÁLISIS DISCRIMINANTE CLÁSICO Y DOS MODELOS DE DISCRIMINACIÓN CON ENFOQUE DEA.

JULIE KIMBERLY RAMIREZ BRIÑEZ

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE CIENCIAS EMPRESARIALES
MAESTRÍA EN INVESTIGACIÓN DE OPERACIONES Y ESTADÍSTICA
PEREIRA
2018

ESTUDIO COMPARATIVO ENTRE UN MODELO DE ANÁLISIS DISCRIMINANTE CLÁSICO Y DOS MODELOS DE DISCRIMINACIÓN CON ENFOQUE DEA.

JULIE KIMBERLY RAMIREZ BRIÑEZ

TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE MAGISTER EN INVESTIGACIÓN DE OPERACIONES Y ESTADÍSTICA.

DIRECTOR:
PhD. JOSÉ ADALBERTO SOTO MEJÍA.

UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE CIENCIAS EMPRESARIALES
MAESTRÍA EN INVESTIGACIÓN DE OPERACIONES Y ESTADÍSTICA
PEREIRA
2018

Nota de aceptación.

_____
_____
_____
_____
_____
_____

_____
Firma de Presidente del Jurado

_____
Firma del Jurado

_____
Firma del Jurado

Pereira, _____del _____ 2018

DEDICATORIA

Dedico ésta tesis a Dios
a mis bellos hijos Nicolás y Tais,
A mi amado esposo Jesús,
a mi hermosa Madre Florelia
y a mi Padre Víctor que está en el cielo.

AGRADECIMIENTOS

A Dios por todas sus bendiciones.
A mi esposo Jesús Ávila por sus muy oportunas voces de aliento y todo su amor.
A mi orientador de tesis PhD. José Adalberto Soto Mejía, por su inconmensurable paciencia.
A la Universidad Tecnológica de Pereira y la Facultad de Ciencias Empresariales.
A los profesores del programa de Maestría en Investigación de Operaciones y Estadística de la Universidad Tecnológica de Pereira.
Y por último, no siendo menos importante a Karol Viviana, secretaria de la maestría por su gran amabilidad en todo este tiempo.

CONTENIDO

Pág.

LISTA DE FIGURAS

LISTA DE TABLAS

RESUMEN

En este trabajo se evalúan tres técnicas de análisis Discriminante: Regla de discriminación vía máxima verosimilitud, modelo de clasificación con solución aproximada por técnica de optimización y modelo DEA-DA extendido. Se realiza la comparación a nivel de dos grupos de identificación y se estima la tasa de error de clasificación por el método de Partición de la muestra. Se usaron dos conjuntos de datos que hacen referencia a información de Empresas y características de consumo e información financiera de bancos japoneses.

Palabras claves: DEA, Análisis Discriminante, Tasa de error de clasificación.

INTRODUCCIÓN

El Análisis Discriminante (AD) forma parte del conjunto de técnicas estadísticas diseñadas para resolver el problema de clasificación o la predicción a la pertenencia de un grupo de una nueva observación. Para hacer uso del enfoque estadístico, un grupo de observaciones, cuya pertenencia a un grupo ya es identificada, son usadas para medir los pesos o estimar los parámetros de las funciones discriminantes, ya sea para minimizar la mala clasificación o maximizar la correcta clasificación. Una nueva observación es clasificada dentro de un grupo, comparando su puntaje discriminante estimado con un puntaje limite, derivado de una función discriminante estimada.

Este trabajo compara el desempeño del modelo clásico de discriminación desde la perspectiva estadística, es decir, usando las funciones lineales y cuadráticas de Fisher, con el desempeño de dos modelos de discriminación con enfoque DEA. Uno de los modelos es el propuesto por Sueyoshi en [17], referido como DEA-DA Extended y el otro modelo propuesto por Almeida en [4], referido como modelo de clasificación con solución aproximada por técnica de optimización.
Se manejan dos conjuntos de datos, usados por los mismos autores y se mide una sola tasa de error de clasificación por el método de Partición de la Muestra, donde el conjunto de datos es dividido en conjunto de calibración y de validación.

El trabajo se divide en 9 capítulos. El capítulo 1 aborda los conceptos básicos y pruebas de la estadística multivariada tomadas de [5], con el fin de conocer la estructura de los datos usados en este estudio, en lo que tiene que ver con: a) cada grupo sigue una distribución normal multivariada, conocido como el supuesto de Normalidad, b) las matrices de covarianzas de cada grupo son iguales, conocido como el supuesto de homogeneidad de varianzas y c) las medias de los vectores y las matrices de covarianzas son conocidas.
En el capítulo 2 se aborda los aspectos generales del Análisis Discriminante, su definición, objetivos, modelos y las diferentes tasas de error de clasificación.
En el capítulo 3 se describe los aspectos teóricos más relevantes del Análisis Envolvente de Datos.
En el capítulo 4 se realiza una descripción un poco más detallada sobre cada uno de los modelos a comparar dentro de este trabajo. En la sección 4.1 se describe un método clásico de discriminación, siguiendo a [3], [5], [6], [10], [11] y [27], en la sección 4.2 el modelo de clasificación con solución aproximada por técnica de optimización propuesto por Almeida en [4] y en la sección 4.3 se aborda el enfoque DEA – DA Extendido, propuesto por Sueyoshi en [16] y [17].
En el capítulo 5 se describen detalladamente las características de los conjuntos de datos a usar en este trabajo.
En el capítulo 6 se explica la metodología de comparación de los tres modelos discriminantes
En el capítulo 7 se interpretan los resultados y se hace un resumen en cuadros comparativos.

Finalmente los capítulos 8 y 9, se realizan las conclusiones y recomendaciones para futuros trabajos.

OBJETIVOS

OBJETIVO GENERAL

Comparar el desempeño entre los siguientes modelos para el análisis discriminante: El enfoque DEA - DA extendido, propuestos por Sueyoshi en [16] y [17], contra el modelo de clasificación con solución aproximada por técnica de optimización propuesto por Almeida en [4] y el modelo clásico de discriminación, siguiendo a [3], [5], [6], [10], [11] y [27]. La comparación se realizará al nivel de clasificación para dos grupos, utilizando conjuntos de datos reales propuestos por los mismos autores en sus diferentes artículos.

OBJETIVOS ESPECIFICOS

- Establecer antecedentes sobre los modelos clásicos que abordan el problema de la discriminación de objetos para dos grupos.
- Realizar una revisión detallada, lo cual incluye la reproducción de resultados, del modelo DEA-DA extendido en [16] y [17] y el modelo de solución aproximada por técnica de optimización propuesta en [4] por Almeida.
- Establecer metodología para comparar los tres modelos antes mencionados, empleando conjunto de datos propuestos por los mismos autores en sus respectivos artículos.
- Medir el desempeño de los tres modelos, usando diferentes metodologías para la estimación de las tasas de clasificación incorrecta.
- Analizar el desempeño de las tres metodologías en aspectos como, supuestos de normalidad y homogeneidad de varianzas y manejo del traslapo.

JUSTIFICACIÓN

En la actualidad el análisis discriminante es empleado en diversas áreas en donde la clasificación de objetos u observaciones es de gran importancia. Algunas de estas aplicaciones mencionadas en [5] son:

- En psicología clínica, donde por ejemplo se desea conformar distintos grupos de alcohólicos, de acuerdo a ciertos tipos de comportamiento.
- En taxonomía vegetal, para la clasificación de especies.
- En estadística inferencial, para muestreo de conglomerados.
- En administración, cuando se trata de ubicar una persona en algunos de los departamentos de una empresa.
- En mercadeo, cuando se busca caracterizar el perfil de los compradores de un determinado producto en un determinado establecimiento o valorar de qué depende la fidelidad de unos clientes a un determinado proveedor comercial
- En finanzas, cuando una entidad bancaria busca clasificar a nuevos clientes como morosos y no morosos.

Según [22] la diversificación en el uso del análisis discriminante está dada por el hecho de que en los últimos años se ha producido un importante crecimiento de las bases de datos en todas las áreas del conocimiento humano. Este incremento es debido principalmente al progreso en las tecnologías para la adquisición y almacenamiento de los datos. Teóricamente, el tener más atributos daría más poder discriminatorio. Sin embargo, la experiencia con algoritmos de aprendizaje ha demostrado que no es siempre así, detectándose algunos problemas: tiempos de ejecución muy elevados, aparición de muchos atributos redundantes y/o irrelevantes, la degradación en el error de predicción, entre otros.

Como el campo de acción del análisis discriminante es bastante amplio, así mismo los investigadores desarrollaron sus propias técnicas de clasificación, según la necesidad y naturaleza de sus propios datos. Ante el incremento de los métodos discriminatorios, este proyecto busca comparar básicamente tres modelos de discriminación al nivel de clasificación de dos grupos y compararlos con el fin de poder establecer cuáles métodos ofrecen mejores resultados, es decir, que ofrezcan un mínimo error de mala clasificación y cuáles se comportan mejor ante ciertas situaciones o estructura de los datos (Normalidad, homogeneidad, etc.). Es aquí donde se deriva la importancia de tener en cuenta nuevos modelos de clasificación, como los que se van a comparar en este trabajo, ya que ante nuevas situaciones mejoran los resultados obtenidos por otras técnicas ya existentes.

ANTECEDENTES

En esfuerzos de investigaciones previas en el Análisis Discriminante (AD), encontramos la primera contribución, desde la perspectiva estadística, realizada por Fisher (1936) y Smith (1947). Este convencional enfoque estadístico del AD, usualmente asume unos supuestos subyacentes (normalidad, homogeneidad varianzas,…) sobre un grupo. Estos enfoques estadísticos otorgan una base teórica, para la realización de varias pruebas estadísticas, lo que puede representar un problema, dado que muchos conjuntos de datos reales, no satisfacen tales supuestos subyacentes.

Para superar tal inconveniente sobre los modelos AD estadísticos, una formulación de la Programación matemática (MP), ha sido propuesta para resolver varios problemas derivados del AD desde el enfoque estadístico. Estas aproximaciones consisten de un grupo de modelos debido a Charnes (1955), el cual documenta cómo formular regresión métrica L_1 con el modelo de programación por metas (GP) y cómo resolver el problema con el algoritmo L_P

Muchos investigadores han puesto especial atención en el Análisis Discriminante basado en Programación por metas, mencionamos los esfuerzos más relevantes:

- Fred y Glover (1981) Maximizar la desviación mínima (MMiD)
- Fred y Glover (1986) Minimizar Desviación Máxima (MMaD)
- Fred y Glover (1986) Minimizar la suma de las desviaciones (MSD)
- Banks y Abad (1991) Minimizar observaciones mal clasificadas (MMO)

Sueyoshi en 1999 [16] propone un nuevo tipo de Análisis discriminante (AD) desde la visión de la programación por metas (GP), esta versión tuvo varios inconvenientes, por ejemplo no podía trabajar con valores negativos en los datos, consecuentemente su aplicación fue muy limitada sobre un conjunto de datos, que comprendieron solo observaciones no negativas.

De esta forma Sueyoshi en 2001 [17], extendió el enfoque propuesto, al trabajar con valores negativos en el conjunto de datos. El enfoque refinado fue referido como "DEA-DA Extendido". Una importante función del enfoque GP extendido, es que este es diseñado para minimizar la distancia total de observaciones mal clasificadas. Además el enfoque propuesto es formulado por formulaciones del GP en dos etapas y puede ser resuelto por algún software de programación lineal. La primera etapa es usada para identificar la existencia de un traslapo entre dos grupos de observaciones. La segunda etapa clasifica observaciones que pertenecen al traslapo.

Para evaluar el desempeño del modelo DEA-DA Extendido, éste ha sido comparado con las funciones lineales y cuadráticas de Fisher, pero no se ha reportado en la literatura su desempeño con el modelo de Almeida, el cual en [4] solo ha sido comparado con las funciones lineales de Fisher, pero no con las funciones cuadráticas de Smith.

ESTADO DEL ARTE

Luego de proponer el modelo DEA-DA extendido en el 2001, Sueyoshi propuso el Enfoque de Programación Entera Mixta del DEA-DA extendido o Enfoque MIP en dos etapas, en el 2004.

En el 2005 Sueyoshi compara el Enfoque MIP en dos etapas, con el MIP Estándar propuesto por Glen en el 2001, para el caso de dos grupos.

En el 2006 Nuevamente Sueyoshi realiza una comparación entre ocho métodos del análisis discriminante, los cuales fueron: Enfoque Estándar MIP, Enfoque MIP en dos etapas, Logit, Probit, Función lineal de Fisher, Función cuadrática de Smith, Red Neuronal y árboles de decisión.

En el 2009 Realizó una comparación entre las industrias japonesas y equipos eléctricos, utilizando el análisis discriminante y en el mismo año realizó una comparación entre el DEA y el DEA-DA desde la perspectiva de la bancarrota financiera.

En el 2011 realizó un estudio del DEA-DA con fuerte condición de holgura complementaria.

Actualmente los trabajos de Sueyoshi no se encuentran enfocados en el Análisis Discriminante, sino en la aplicación de los modelos DEA para mejorar la sostenibilidad ambiental.

Sueyoshi es citado en varios trabajos, por ejemplo Stern y Friedman (1998) en su artículo "DEA and the discriminant analysis of ratios for ranking Units", Balf, Saghaei y Lotfi (2010) en su artículo "Análisis Discriminante para tres grupos", Hasan Bal y Hassan Orkcu (2011) en su artículo "Un nuevo enfoque de programación matemática para los problemas de clasificación multigrupo"

1. CONCEPTOS BÁSICOS DE ESTADÍSTICA MULTIVARIADA

Dentro de las metodologías para el análisis de datos multivariados, encontramos los métodos de dependencia y de interdependencia. En los métodos de dependencia se suponen que las variables analizadas, están divididas en dos grupos: las variables dependientes y las variables independientes y el objetivo principal consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma. Las técnicas de análisis de interdependencia buscan el cómo y el por qué se relacionan o asocian un conjunto de variables.

El análisis discriminante de datos hace parte de los métodos de dependencia junto con la regresión simple, el análisis de correlación canónica, análisis logit, análisis de varianza multivariado y el análisis conjunto.

En este capítulo se presenta los conceptos básicos y pruebas de la estadística multivariada tomadas de [5], con el fin de conocer la estructura de los datos que se usaran en este estudio, en lo que tiene que ver con: a) cada grupo sigue una distribución normal multivariada, conocido como el supuesto de Normalidad, b) las matrices de covarianzas de cada grupo son iguales, conocido como el supuesto de homogeneidad de varianzas y c) las medias de los vectores y las matrices de covarianzas son conocidas.

El desempeño de un modelo puede verse afectado o no, incluso en muchos casos puede aumentar el buen desempeño, dependiendo de la estructura de los grupos en cada conjunto de datos.

1.1 MATRIZ DE VARIANZAS Y COVARIANZAS

Según [5], dado un vector aleatorio X, la matriz de varianzas y covarianzas de X, la cual notaremos por Σ, está dada por:

$$\boldsymbol{\Sigma} = Cov(\boldsymbol{X}) = \varepsilon\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{1.1}$$

Donde σ_{ij} denota la covarianza entre la variable X_i y la variable X_j, la cual se define como:

$$\sigma_{ij} = \varepsilon[(X_i - \mu_i)(X_j - \mu_j)] \tag{1.2}$$

Como propiedades de la matriz de varianzas y covarianzas tenemos:

- La matriz $\boldsymbol{\Sigma}$ es simétrica; es decir, $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$ puesto que $\sigma_{ij} = \sigma_{ji}$
- Los elementos de la diagonal de $\boldsymbol{\Sigma}$ corresponden a la varianza de las respectivas variables ($\sigma_{ii} = \sigma_i^2$), los elementos fuera de la diagonal, son las covarianzas entre las variables correspondientes de la fila y la columna.

- Toda matriz de varianzas y covarianzas es definida no negativa ($|\boldsymbol{\Sigma}| \geq 0$) y es definida positiva cuando el vector aleatorio es continuo.
- Si $\varepsilon(\boldsymbol{X}) = \mu$ y $Cov(\boldsymbol{X}) = \boldsymbol{\Sigma}$, entonces:
  $\varepsilon(\boldsymbol{A'X} + \boldsymbol{b}) = \boldsymbol{A\mu} + \boldsymbol{b}$ y $Cov(\boldsymbol{A'X} + \boldsymbol{b}) = \boldsymbol{A\Sigma A'}$, con $\boldsymbol{A}$ matriz de constantes de tamaño $(q \times p)$ y $\boldsymbol{b}$ vector $(q \times 1)$ también de constantes.

A continuación se desarrollan algunas estadísticas descriptivas ligadas a los parámetros anteriores. Sea x_{ij} la observación de la $j-$ésima variable en el $i-$ésimo individuo, se define la matriz de datos multivariado como el arreglo

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{1.3}$$

El vector formado por las *p*-medias muestrales, es el vector de promedios o de medias (centroide de los datos)

$$\bar{\mathbf{X}}' = \frac{1}{n}\mathbf{1}'\mathbf{X} = (\bar{x}_1, \ldots, \bar{x}_p) \tag{1.4}$$

Donde **1'** es el vector columna de unos. Se define la media muestral de la $j-$ésima variable por

$\bar{x}_j = \sum_{i=1}^{n} x_{ij}$

La matriz constituida por las covarianzas s_{ij}, es la matriz de varianzas y covarianzas muestral, ésta es:

$$\boldsymbol{S} = \frac{1}{n-1}\mathbf{X}'\left(\boldsymbol{I_n} - \frac{1}{n}\mathbf{1}'\mathbf{1}\right)\mathbf{X}' = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & & s_{2p} \\ \vdots & & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \tag{1.5}$$

1.2 LA DISTRIBUCIÓN NORMAL MULTIVARIADA

Sea $Z' = (Z_1, \ldots, Z_p)$ un vector con p variables aleatorias independientes y cada una con distribución normal estándar; es decir $Z_i \sim n(0,1)$. Entonces

$\varepsilon(Z) = 0$, $Cov(Z) = \boldsymbol{I}$, $M_Z(t) = \prod_{i=1}^{p}\left\{\frac{t_i^2}{2}\right\} = exp\frac{t\prime t}{2}$

Considérese el vector $\boldsymbol{\mu}$ y la matriz A de tamaño $(p \times p)$. El vector $\mathrm{X} = AZ + \boldsymbol{\mu}$ es tal que

$\varepsilon(\mathrm{X}) = \boldsymbol{\mu}$, $Cov(\mathrm{X}) = AA'$

El vector p-dimensional X, tiene distribución normal p-variante, con vector de medias μ y matriz de covarianzas Σ, si y sólo si, la función generadora de momentos de X es:

$$M_X(t) = \exp\left\{\mu' t + \frac{t'(\Sigma)t}{2}\right\}$$

Se nota $\boldsymbol{X} \sim N_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Ahora se encuentra la función de densidad para X. Del resultado anterior se afirma que

$$Z \sim N_P(0, \boldsymbol{I}), \quad con \;\; Z' = (Z_1, \dots, Z_P)$$

Por la independencia entre los Z_i su densidad conjunta es,

$$f_Z(z) = \prod_{i=1}^{p} \frac{1}{(2\pi)^{1/2}} exp\left\{-\frac{1}{2} z_i^2\right\} = \frac{1}{(2\pi)^{p/2}} exp\left\{-\frac{1}{2} z'z\right\}$$

Sea $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$, entonces por el resultado anterior $\boldsymbol{X} \sim \boldsymbol{N}_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. El vector **Z** se puede expresar como $\boldsymbol{Z} = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$, expresión que es una transformación invertible. El jacobiano de la transformación es $\boldsymbol{J} = |\boldsymbol{\Sigma}|^{1/2}$, por tanto la función de densidad conjunta de **X** es

$$f_X(x) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(x - \boldsymbol{\mu})'\Sigma^{-1}(x - \boldsymbol{\mu})\right\} \qquad (1.6)$$

Donde $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ y **Σ** es una matriz simétrica definida positiva de tamaño $p \times p$.

1.3 PRUEBAS ESTADISTICAS

1.3.1 Prueba de Multinormalidad de Mardia.

Para contrastar Multinormalidad no es suficiente con probar la normalidad de las distribuciones marginales, puesto que se estaría dejando de lado la asociación lineal entre las variables, la cual se refleja a través de la matriz de covarianzas. La idea para contrastar Multinormalidad es una extensión de alguna de las pruebas univariadas.

Mardia (1970) define los coeficientes de simetría y curtosis multivariados, para un vector $\boldsymbol{X}$ de tamaño $(p \times 1)$ con media μ y matriz de dispersión Σ, mediante la siguiente expresión:

$$\beta_{1,p} = \varepsilon[\{(X - \mu)'\Sigma^{-1}(Y - \mu)\}^3]$$
$$\beta_{2,p} = \varepsilon[\{(X - \mu)'\Sigma^{-1}(X - \mu)\}^2] \qquad (1.7)$$

Donde X y Y son independientes e idénticamente distribuidos. Estas medidas son invariantes por transformaciones lineales. Si $X \sim N_P(\mu, \Sigma)$, entonces, los coeficientes de simetría y curtosis son, respectivamente, $\beta_{1,p} = 0$ y $\beta_{2,p} = p(p + 2)$.

La generalización de las medidas se observa porque $\sqrt{\beta_{1,1}} = \sqrt{\beta_1}$ y $\beta_{2,1} = \beta_2$. Se puede contrastar las hipótesis sobre estos valores empleando los siguientes estimadores muestrales

$b_{1.p} = \frac{1}{n^2}\sum_{h=1}^{n}\sum_{i=1}^{n} g_{hi}^3$ y $b_{2,p} = \frac{1}{n}\sum_{i=1}^{n} g_{ii}^2$, con
$g_{hi} = (x_h - \bar{x})\left(\frac{A}{n}\right)^{-1}(x_i - \bar{x})$, y $A = \sum_i (x_i - \bar{x})(x_i - \bar{x})'$ (1.8)

Mardia (1970) demuestra que bajo la hipótesis de distribución normal Multivariante, se tiene la distribución asintótica de

$B_1 = \frac{n}{6} b_{1,p} \sim \chi_f^2$, donde $f = \frac{1}{6}p(p+1)(p+2)$ (1.9)
Se rechaza la hipótesis de simetría entorno a la media, $\left(H_0: \sqrt{\beta_{1,1}} = 0\right)$ si $B_1 \geq \chi_{\alpha,f}^2$.
Para verificar que el coeficiente de curtosis no es significativamente diferente de $p(p+2)$ se emplea la estadística

$B_2 = \frac{b_{2,p} - p(p+2)}{[8p(p+2)/n]^{1/2}} \sim n(0,1)$ (1.10)

1.3.2 Prueba de homogeneidad de varianzas.

La igualdad de matrices de covarianzas es un supuesto que se requiere para aplicar adecuadamente algunas técnicas tales como la comparación de medias en dos o más poblaciones, el análisis discriminante, entre otras (Morrison, 1990, pág., 293).

En el caso multivariado se trata de contrastar la hipótesis sobre la igualdad de las matrices de covarianzas asociadas a varias poblaciones multinormales, mediante la información contenida en una muestra aleatoria de cada una de ellas.
Sea $X_{1g}, \dots, X_{ng}$, con $g = 1, \dots, q$, una muestra aleatoria de una población $N_p(\mu_g, \Sigma_g)$; es decir, se dispone de q muestras independientes de poblaciones multinormales, la hipótesis a contrastar es

$H_0 = \Sigma_1 = \cdots = \Sigma_q = \Sigma$ (1.11)

De los datos muestrales se obtiene las matrices

$A_g = \sum_{\alpha=1}^{n_g}\left(X_{\alpha g} - \bar{X}_g\right)\left(X_{\alpha g} - \bar{X}_g\right)'$
$A = \sum_{g=1}^{q} A_g$
$\sum_{g=1}^{q} n_g = N$, con $g = 1, \dots, q$

Mediante las matrices A_g y A se estiman Σ_g y Σ, en el espacio de parámetros general y en el espacio de parámetros reducidos por H_0, respectivamente. Así,

$\hat{\Sigma}_g = \frac{1}{n_g} A_g$ y $\hat{\Sigma} = \frac{1}{N} A$

Considerando $v_g = (n_g - 1)$ y $v = \sum_{q=1}^{q} v_g = (N - q)$, se obtienen los estimadores insesgados para Σ_g y Σ; estos son respectivamente S_p y S_g; es decir:

$$S_g = \frac{1}{v_g} A_g \text{ y } S_p = \frac{1}{v} A = \frac{1}{v} \sum_{g=1}^{q} v_g S_g \tag{1.12}$$

La razón de máxima verosimilitud para verificar (1.11) es:

$$\lambda_1 = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2} n_g}}{|A|^{\frac{1}{2} N}} \frac{n^{\frac{1}{2} pN}}{\prod_{g=1}^{q} n_g^{\frac{1}{2} p n_g}} \tag{1.13}$$

Se rechaza H_0 para valores pequeños de λ_1 a un nivel de significación α; es decir, se rechaza H_0 para valores λ_1 tales que $\lambda_1 \leq \lambda_1(\alpha)$
Una modificación de (1.12) fue propuesta por Bartlett (1937) para el caso univariado $(p = 1)$, donde se reemplazan los tamaños muestrales por los grados de libertad de A_g y de A; esto es n_g por $v_g = (n_g - 1)$ y N por $v = \sum_{g=1}^{q} v_g = (N - q)$. La estadística correspondiente equivalente con la anterior estadística es:

$$\lambda_1 = \frac{\prod_{g=1}^{q} |A_g|^{\frac{1}{2} v_g}}{|A|^{\frac{1}{2} v}}$$

Para dos muestras $q = 2$ y $p = 1$

$A_1 = \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 = v_1 s_1^2$
$A_2 = \sum_{i=1}^{n_1} (x_{i2} - \bar{x}_2)^2 = v_2 s_2^2$
$A = A_1 + A_2 = v_1 s_1^2 + v_2 s_2^2 = (v_1 + v_2) s_p^2$

Las estadísticas s_1^2 y s_2^2 son los estimadores insesgados de σ_1^2 y σ_2^2 al reemplazarlas en (1.13) resulta

$$\lambda_1 \frac{(v_1)^{\frac{1}{2} v_1} (v_2)^{\frac{1}{2} v_2} (s_1^2)^{\frac{1}{2} v_1} (s_2^2)^{\frac{1}{2} v_2}}{(v_1 s_1^2 + v_2 s_2^2)^{\frac{1}{2}(v_1 + v_2)}}$$

Recuérdese que la estadística s_1^2 / s_2^2 tiene distribución F y se emplea para verificar la hipótesis H_0: $\sigma_1^2 = \sigma_2^2$. Si se divide la última expresión por $(s_2^2)^{\frac{1}{2}(v_1 + v_2)}$ se obtiene

$$\lambda_1 \frac{(v_1)^{\frac{1}{2} v_1} (v_2)^{\frac{1}{2} v_2} F^{\frac{1}{2} v_1}}{(v_1 F + v_2)^{\frac{1}{2}(v_1 + v_2)}}$$

La región crítica está dada por los valores muestrales tales que $\lambda_1 \leq \lambda_1(\alpha)$. La cual es función de $F_{(n_1,n_2)}$. La región crítica queda determinada por los valores de F tales que $F \leq F_1(\alpha)$ o $F \leq F_2(\alpha)$.

La distribución asintótica de λ_1 se obtiene al reemplazar n_g por v_g y N por v. Al aplicar logaritmos en los dos lados de la nueva expresión para λ_1 y sustituir A_g por $n_g s_g$ y A por NS_p, se obtiene

$$-2ln(\lambda_{1n}) = vln|S_p| - \sum_{g=1}^{q} v_g ln\,|S_g|$$

Box (1949) demuestra que si se introduce la cantidad ρ dada por

$$\rho = 1 - \frac{2p^2+3p-1}{6(p+1)(q-1)}\left(\sum_{g=1}^{q}\frac{1}{v_g} - \frac{1}{v}\right),$$

Entonces

$$\varphi = -2\rho \ln(\lambda_{1n})$$

Se distribuye asintóticamente como ji-cuadrado con $p(p+1)(q-1)/2$ grados de libertad.

1.3.3 Prueba de igualdad de medias.

Considérese dos muestras de poblaciones normales p-variantes e independientes. Supóngase que $(X_{\alpha 1})$, es una muestra de tamaño n_1 de una población $N(\mu_1, \Sigma_1)$ y $(X_{\alpha 2})$, es una muestra de tamaño n_2 de una población $N(\mu_2, \Sigma_2)$, con $\alpha_i = 1, \ldots, n_i$ e i=1,2. En estas condiciones la estadística T^2 puede emplearse para contrastar la hipótesis que la media de una población es igual a la media de la otra; donde la matriz de covarianzas, aunque desconocida, se supone igual.

El vector de medias muestral $\bar{X}_i$ tiene distribución $N_p(\mu_i, \frac{1}{n_i}\Sigma)$, para i=1,2. Así, el vector aleatorio $[n_1 n_2/(n_1+n_2)]^{1/2}(\bar{X}_1 - \bar{X}_2)$, se distribuye como $N_p(0, \Sigma)$. La matriz de covarianzas Σ, se estima en forma mancomunada con las matrices de covarianzas muestrales; así,

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

La estadística

$$T^2 = \frac{n_1 n_2}{(n_1+n_2)}(\bar{X}_1 - \bar{X}_2)' S_p^{-1}(\bar{X}_1 - \bar{X}_2)$$

Se distribuye como T^2 con dimensión p y $v = n_1 + n_2 - 2$ grados de libertad. La región crítica para contrastar la hipótesis $H_0: \mu_1 = \mu_2$ es

$$T^2 > \frac{vp}{(v-p+1)} F_{(p,v-p+1)}(\alpha)$$

Con un nivel de significancia igual a α. Una región de confianza para $\mu_1 - \mu_2$, con un nivel de confiabilidad de $(1-\alpha)\%$ es el conjunto de vectores m que satisfacen:

$$\left((\bar{X}_1 - \bar{X}_2) - m\right)' S_p^{-1} ((\bar{X}_1 - \bar{X}_2) - m) \leq \frac{(n_1+n_2)}{n_1 n_2} T_v^2(\alpha)$$

$$\left((\bar{X}_1 - \bar{X}_2) - m\right)' S_p^{-1} ((\bar{X}_1 - \bar{X}_2) - m) = \frac{(n_1+n_2)}{n_1 n_2} \frac{vp}{(v-p+1)} F_{(p,v-p+1)}(\alpha)$$

2. ANÁLISIS DISCRIMINANTE.

En el capítulo 1, se hizo referencia sobre las diferentes metodologías utilizadas para el análisis de datos multivariados, dentro de los cuales encontramos los métodos de dependencia y de interdependencia. Específicamente en los métodos de dependencia encontramos el análisis discriminante; en éste capítulo se aborda su definición, objetivos, modelos y las diferentes tasas de error de clasificación.

2.1 DEFINICIÓN Y OBJETIVOS DEL ANÁLISIS DISCRIMINANTE.

El análisis discriminante es una técnica estadística que permite asignar o clasificar a distintos individuos dentro de grupos previamente reconocidos o definidos [9] y [27]. Cada individuo puede pertenecer a un solo grupo. La pertenencia a uno u otro grupo se introduce en el análisis mediante una variable categórica que toma tanto valores como grupos existentes. En el análisis discriminante esta variable juega el papel de variable dependiente, mientras que las variables que se utilizan para realizar la clasificación de los individuos se denominan variables clasificadoras o explicativas. Varios son los objetivos principales abordados por el análisis discriminante [3], [5], [6]:

- Trata de determinar la contribución de cada variable clasificadora a la clasificación correcta de cada uno de los individuos (fin explicativo o descriptivo). Esto implica, encontrar reglas que expliquen el agrupamiento de los objetos u observaciones y describir las diferencias entre grupos, para posteriormente realizar la exploración de los grupos conformados, en otras palabras, trata de encontrar las diferencias entre dos o más grupos a través de una función discriminante.
- Determina el grupo al que pertenece un nuevo individuo para el que se conocen los valores que toman las variables clasificadoras (fin predictivo), frecuentemente la mejor función para separar grupos provee también la mejor regla de localización de observaciones futuras; de tal forma que estos dos términos generalmente se emplean indistintamente.
- Debe encontrar un subconjunto de variables que den mayor poder de separación y decidir cuáles variables difieren en los grupos.

2.2 TASAS DE ERROR DE CLASIFICACIÓN

En [5] una vez que se ha obtenido una regla de clasificación por medio de alguna técnica, la inquietud más importante y natural es acerca de qué tan buena es la clasificación generada a través de esta regla. Es decir, se quiere conocer *la tasa de clasificación correcta*, referida como la probabilidad de clasificar una observación en el grupo al que verdaderamente pertenece. De manera complementaria se tienen las tasas de error por clasificación incorrecta.

Grupo	Asignar a G_1	Asignar a G_2
$G_1(n_1)$	Decisión correcta (n_{11})	Error (n_{12})
$G_2(n_2)$	Error (n_{21})	Decisión correcta (n_{22})

Tabla 1: Calidad de las decisiones Estadísticas en relación a la clasificación de objetos.

Existe un método simple que permiten estimar la tasa de error de clasificación conocido como *Resustitución*. A cada observación se le aplica la función de clasificación y se asigna a uno de los grupos, se cuenta entonces el número de clasificaciones correctas y el número de clasificaciones incorrectas conseguidas con la regla. La proporción de clasificaciones incorrectas se conoce como *tasa de error aparente* y dicho método se puede extender al caso de clasificación multigrupo. Para muestras de tamaño grande la *tasa de error aparente* como un estimador de la tasa de error tiene un sesgo pequeño. Para muestras de tamaño pequeño, la situación respecto a la disminución del sesgo no es muy buena. Algunas técnicas que permiten reducir el sesgo en la estimación de la tasa de error aparente son:

2.2.1 Partición de la muestra.

El cual es una forma de controlar el sesgo mediante la división de la muestra en dos partes. Una de ellas es la muestra de ensayo la cual se emplea para construir la regla de clasificación mientras que la otra llamada muestra de validación se utiliza para evaluar la bondad de la regla calculada.

2.2.2 Validación cruzada.

Se considera como un caso especial del anterior pues se toman $n-1$ observaciones para construir la regla de clasificación y luego con ella se clasifica la observación omitida. Este procedimiento se repite una vez por cada observación. Esta técnica es llamada también Leave One Out (LOO), dejar uno por fuera.

2.2.3 Estimación Bootstrap.

Es esencialmente una corrección del sesgo para la tasa de error aparente basados sobre un remuestreo de la muestra original. Se describe el procedimiento para dos grupos con tamaños de muestra diferente n_1 y n_2, donde en la primera muestra se toma una muestra aleatoria de tamaño n_1 **con reemplazamiento** y de manera similar se remuestrea el segundo grupo. Con las dos nuevas muestras se recalculan las funciones de clasificación y con estas se clasifican tanto las muestras originales como las nuevas.
Las tasas de error en la clasificación para cada grupo se calculan con

$$d_i = \frac{e_{i.orig} - e_{i.nvo}}{n-i}, \quad i = 1,2;$$

Donde $e_{i.orig}$ es el número de observaciones del i-ésimo grupo original incorrectamente clasificadas y $e_{i.nvo}$ es el número de observaciones de la i-ésima muestra nueva que fueron mal clasificados. Este procedimiento se desarrolla un buen número de veces (se sugieren

entre 100 y 200 repeticiones) y se emplea el promedio de los d_i como corrector del término del sesgo, así:

$$Tasa\ de\ error\ bootstrap = tasa\ de\ error\ aparente + \bar{d}_1 + \bar{d}_2$$

2.3 MODELOS DE DISCRIMINACIÓN

Según en [18], las estrategias más conocidas y utilizadas para la clasificación de objetos u observaciones en dos o más grupos son: en el área de estadística multivariada, con los métodos de análisis de conglomerados y análisis discriminante; en programación matemática, con los métodos de programación lineal, no lineal y entera mixta; estos modelos serán expuestos ampliamente en el capítulo 4. Otras estrategias para la clasificación de objetos son las econométricas con los modelos de discriminación logit y probit, las metaheurísticas con métodos de algoritmos genéticos y redes neuronales; los cuales se exponen brevemente a continuación para el caso de dos grupos.

2.3.1 Modelo de discriminación logística.

En [5], cuando las variables son discretas o son una mezcla de discretas y continuas, la discriminación a través del modelo logístico puede resultar adecuada.

Para distribuciones multinormales con $\Sigma_1 = \Sigma_2 = \Sigma$, el logaritmo de la razón de densidades es

$$ln\frac{f(X/G_1)}{f(X/G_2)} = -\frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) + ((\mu_1 - \mu_2)'\Sigma^{-1}X$$

$$ln\frac{f(X/G_1)}{f(X/G_2)} = \alpha + \beta'X \qquad (2.1)$$

La cual es una función lineal del vector observado X. Además de la normal multivariada, otras distribuciones multivariadas satisfacen (2.1), algunas de las cuales involucran vectores aleatorios discretos o mezclas de variables discretas y continuas. El modelo mostrado se conoce como logístico. La regla para ubicar una observación X es: Asignar al grupo G_1 si:

$$\alpha + \beta'X > ln\frac{p_1}{p_2} \qquad (2.2)$$

y a G_2 en otro caso. Cuando las probabilidades a priori, p_1 y p_2, se pueden asumir iguales, el miembro izquierdo de la desigualdad (2.2) se compara contra el número cero. La clasificación logística es también referida como la discriminación logística.

2.3.2 Modelo de discriminación probit.

En algunos casos los grupos son definidos a través de un criterio cuantitativo en lugar de cualitativo.

Sea Z una variable aleatoria continua, si t es un valor "umbral" entonces un individuo es asignado al grupo G_1 si $Z > t$ y si $Z \leq t$ se asigna al grupo G_2.

Se asume que el vector $(Z, X)'$ se distribuye $N_{p+1}(\mu, \Sigma)$, la distribución condicional de Z dado el vector X es normal con

$$\varepsilon(Z/X) = \mu_{Z/X} = \mu_Z + \sigma_{ZX}\Sigma_{XX}^{-1}(X - \mu_X),$$

$$var(Z/X) = \sigma_{ZX} = \sigma_Z^2 - \sigma_{ZX}\Sigma_{XX}^{-1}\sigma_{XZ}$$

Por tanto

$$P(G_1/X) = \Phi\left(\frac{-t+\mu_{Z/X}}{\sigma_{Z/X}}\right)$$

Donde $\Phi(.)$ es la función de distribución normal estándar. De esta forma reemplazando por las expresiones anteriores $\mu_{Z/X}$ y $\sigma_{Z/X}$ la probabilidad de que la observación X sea del grupo G_1 es de

$$P(G_1/X) = \Phi\left(\frac{-t+\mu_Z+\sigma_{ZX}\Sigma_{XX}^{-1}(X-\mu_X)}{\sqrt{\sigma_Z^2-\sigma_{ZX}\Sigma_{XX}^{-1}\sigma_{XZ}}}\right) = \Phi(\gamma_0 + \gamma_1 X)$$

Donde

$$\gamma_0 = -(t - \mu_Z + \sigma_{ZX}\Sigma_{XX}^{-1}(X - \mu_X)/\sqrt{\sigma_Z^2 - \sigma_{ZX}\Sigma_{XX}^{-1}\sigma_{XZ}} \text{ y}$$

$$\gamma_1 = \sigma_{ZX}\Sigma_{XX}^{-1}\sqrt{\sigma_Z^2 - \sigma_{ZX}\Sigma_{XX}^{-1}\sigma_{XZ}}$$

La regla de clasificación asigna la observación X al grupo G_1 si

$$P(Z > t/X) \geq P(Z < t/X)$$

Es decir, si $P(G_1/X) \geq P(G_2/X)$, y al grupo G_2 en otro caso, de esta forma la regla es: Asignar la observación X al grupo G_1 si $\Phi(\gamma_0 + \gamma_1 X) \geq 1 - \Phi(\gamma_0 + \gamma_1 X)$ lo cual equivale a que $\Phi(\gamma_0 + \gamma_1 X) \geq 1/2$. En términos de $\gamma_0 + \gamma_1 X$, la regla puede expresarse como: asignar a X al grupo G_1 si

$$\gamma_0 + \gamma_1 X \geq 0$$

Y al grupo G_2 en otro caso, los parámetros γ_0 y γ_1 se estiman a través del método de máxima verosimilitud (con soluciones iterativas) empleando una dicotomización del tipo

$\omega = 0$ si $Z \leq t$ y $\omega = 1$ si $Z > t$. No se requiere que X tenga una distribución multinormal, únicamente que la distribución condicional de Z dado X sea normal. Esto posibilita la inclusión en X de variables aleatorias discretas.

2.3.3 Clasificación mediante la técnica del "Vecino más cercano"

El método de clasificación llamado "el vecino más cercano", se considera como una técnica de tipo no paramétrico. Para el procedimiento se determina la distancia de Mahalanobis de una observación X_i, respecto a las demás observaciones X_j mediante

$$D_{ij} = \left(X_i - X_j\right)'^{S_p^{-1}}\left(X_i - X_j\right),\ i \neq j$$

Para clasificar la observación X_i en uno de los grupos se examinan los k puntos más cercanos a X_i, si la mayoría de estos k puntos pertenecen al grupo G_1 se asigna la observación X_i a G_1, en otro caso se asigna a G_2. Si se nota el número de individuos (objetos) de G_1 por k_1 y a los restantes por k_2 en G_2, con $k = k_1 + k_2$, entonces la regla se expresa también como: asignar X_i a G_1 si

$k_1 > k_2$

Y G_2 en otro caso. Si los tamaños muestrales de cada grupo son n_1 y n_2 respectivamente, la decisión es: asignar X_i a G_1 si

$\frac{k_1}{n_1} > \frac{k_2}{n_2}$

De una manera coloquial, una observación X_i se asigna al grupo donde se inclinen la mayoría de sus vecinos; es decir, por votación la mayoría decide el grupo donde se debe ubicar cada observación.

Además si se consideran las probabilidades a priori: asignar x_i a G_1 si

$\frac{k_1/n_1}{k_2/n_2} > \frac{p_1}{p_2}$

2.3.4 Clasificación mediante redes neuronales

Se ha observado que muchos problemas en patrones de reconocimiento han sido resueltos más fácilmente por humanos que por computadores, tal vez por la arquitectura básica y el funcionamiento de su cerebro. Las redes neuronales son diseñadas mediante emulaciones, hasta ahora incompletas con el cerebro humano para imitar el trabajo humano y tal vez su inteligencia. El término red neuronal artificial es usado para referirse a algoritmos de cómputo que usan las estructuras básicas de las neuronas biológicas.

Una neurona recibe impulsos de otras neuronas a través de las dendritas. Los impulsos que llegan son enviados por los terminales de los axones a las otras neuronas. La transmisión de una señal de una neurona a otra se hace a través de una conexión (sinapsis) con las dendritas de las neuronas vecinas. La sinapsis es un proceso físico químico complejo, el cual genera una inversión de potencial en célula receptora; si el potencial alcanza cierto umbral, la célula envía una señal a través de su axón y en consecuencia se establece una comunicación con las que se le conecten directa e indirectamente.

Una neurona artificial (en adelante una simple neurona) en computación consta: de unas entradas o estímulos, una caja de procesamiento y una respuesta. El modelo más simple de

neurona artificial es el modelo de McCulloch y Pits. Supóngase que la atención está sobre la neurona k, esta neurona recibe una serie de entradas Y_{ik} , cada una de las cuales puede ser la salida de la $i-$ ésima neurona vecina. La neurona desarrolla una suma ponderada de las entradas y produce como salida un cero o un uno dependiendo de si la suma supera un umbral μ_k asignado a la neurona. El gráfico 1. Ilustra este modelo de neurona.

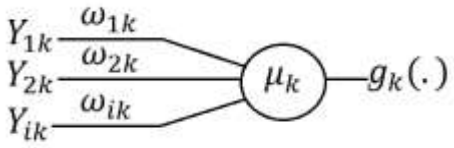

Gráfico 1: Modelo de Neurona Simple

Las entradas $Y_{1k}, \dots, Y_{ik}$ corresponden a las salidas de las neuronas conectadas con la neurona k.
Las cantidades $\omega_{1k}, \dots, \omega_{ik}$ son las ponderaciones de conexión entre la salida de $j-$ ésima neurona y la entrada a la $k-$ ésima neurona.
μ_k Es el umbral de la señal de la $k-$ ésima neurona.
$g_k(.)$ Es la función de salida, respuesta o transferencia de la $k-$ ésima neurona.
La ecuación de nodo es

$$Z_k = g_k\left(\sum_j \omega_{jk} Y_{jk} - \mu_k\right) = \begin{cases} 1, & si \ \sum_j \omega_{jk} Y_{jk} \geq -\mu_k \\ 0, & si \ \sum_j \omega_{jk} Y_{jk} < \mu_k \end{cases}$$

Una red consiste en un conjunto de neuronas o unidades de cómputo. Cada neurona en una red desarrolla un cálculo simple. Tres son los elementos básicos de una red neuronal: las neuronas, nodos o unidades de cómputo; la arquitectura (topología) de una red, la cual describe las conexiones entre los nodos y el algoritmo de entrenamiento, usado para encontrar los valores particulares de los parámetros, con los cuales la red desarrolla eficientemente una tarea particular.
Un perceptrón es una red neuronal, que están conformados por varias neuronas que desarrollan un trabajo específico. Un perceptrón multicapa está constituido por varias capas de neuronas interconectadas con alguna arquitectura específica, este tipo de modelos es el que más atención ha recibido para clasificación.
Rosenblant (1962), demuestra que si dos conjuntos de datos, se separan por un hiperplano, entonces mediante el modelo tipo perceptrón se determina un plano que los separe.
La asignación de un individuo determinado por el vector $X' = (X_1, \dots, X_p)$ a uno de los $q-$grupos $G_1, \dots, G_q$, puede verse como un proceso matemático que transforma las p entradas $X_1, \dots, X_p$ en q unidades de salida $Z_1, \dots, Z_q$ las cuales definen la localización de un individuo en un grupo; es decir, $Z_i = 0$ y $Z_j = 0$ para todo $1 \neq j$ si el individuo es localizado en el grupo G_i. El perceptrón multicapa lleva a cabo la tarea multitransformación tratando a los X_i como valores de $p-$unidades, en la capa de entrada, los Z_j son los valores de las $q-$unidades en la capa de salida; además entre estas dos capas hay algunas capas escondidas (intermedias) de nodos o neuronas. Usualmente cada unidad en una capa está conectada a todas las unidades de la capa adyacente y no a otras (aunque algunas redes permitan conectar unidades de capas no contiguas). La arquitectura

o topología de una red es determinada por el número de capas, el número de unidades en cada capa, y las conexiones entre unidades.

3. ANÁLISIS ENVOLVENTE DE DATOS.

Dos de los modelos de discriminación a comparar en este trabajo incorpora en su estructura el Análisis Envolvente de Datos, en este capítulo se describe los aspectos teóricos más relevantes de esta técnica.

Según [14] El Análisis Envolvente de Datos en inglés Data Envelopment Analysis conocido como DEA, es un método propuesto inicialmente por Charnes, Cooper y Rhodes (1978). Es una herramienta de la investigación de operaciones, basada en la programación lineal en donde se resuelve el problema de maximizar la eficiencia de cada unidad de decisión DMU (Decision Making Units).
Un DMU puede ser una dependencia, un proceso o un grupo que consuma recursos y genere productos.

3.1 MODELO DEA CCR (CHARNES, COOPER. & RHODES)

El cálculo usual de eficiencia, califica como más eficientes aquellas unidades organizacionales que usan de manera intensiva sus recursos (entradas) obteniendo mayores salidas (Productos)

$$Eficiencia = \frac{Productos}{Entradas}$$

Pero esta ecuación es inadecuada, cuando existen múltiples entradas (recursos) y salidas (productos) relacionadas con diferentes recursos, que se expresan en diferentes unidades.

El DEA calcula la eficiencia a partir de la siguiente expresión:

$$h_{j0} = \frac{\sum_r u_r y_{rj0}}{\sum_i v_i x_{ij0}}$$

En donde:

$r = 1, \dots, m$ Subíndice que identifica un producto
$j = 1, \dots, n$ Subíndice que identifica las diferentes unidades de decisión.
j_0 Subíndice que indica la unidad de decisión a la que se le está calculando la eficiencia.
h_{jo} Es la eficiencia de la unidad de decisión que se está calculando.
u_r Es el peso que tiene el producto y_r para la DMU j_0 que se está calculando.
v_i Es el peso que tiene el insumo x_i en la DMU j_0 que está siendo calculada.

La expresión anterior es utilizada como la función objetivo de un modelo de programación lineal que busca maximizar esa eficiencia sujeta a las siguientes restricciones:

$$\frac{\sum_r u_r y_{rj0}}{\sum_i v_i x_{ij0}} \leq 1$$
$$u_r, v_i \geq \varepsilon$$

Las restricciones anteriores garantizan que al calcular la eficiencia de la DMU a la que se le está calculando la eficiencia al variar los pesos, estos no generarán eficiencias mayores que 1. Las variables de decisión son los pesos de los productos y de los insumos u_r, v_i respectivamente. El número ε es un valor de perturbación, que permite que los pesos de todos los productos e insumos puedan ser mayores a cero.

La anterior expresión posee en la función objetivo y en las restricciones expresiones que deben modificarse para que tomen la forma estándar del modelo de programación lineal, por lo que se debe hacer la siguiente manipulación:

Como la función objetivo es una fracción basta con maximizar el numerador (output) bajo un denominador constante para que la expresión alcance su mayor valor o bajo un numerador constante minimizar el denominador (inputs)

Para el primer caso el modelo presentado quedará así:

$\max h_0 = \sum_r u_r y_{rj0}$
Sujeto a:
$\sum_i v_i x_{ij0} = 100$
$\sum_r u_r y_{rj0} - \sum_i v_i x_{ij0} \leq 0$
$u_r, v_i \geq \varepsilon$

La eficiencia de cada unidad de decisión se obtiene cuando se resuelve el modelo de programación lineal que se acaba de presentar. Por lo que se deben resolver tantos modelos como unidades de decisión j existan. La solución al modelo garantiza que se den los mejores pesos u_r y v_i a los productos e insumos respectivamente de acuerdo a la conveniencia de cada DMU. Así al construirse la lista ordenada (Ranking) de los más eficientes, el índice de eficiencia, es el mejor posible y el más conveniente, lo que no permite argumentos en contra como "que los productos y_{rj}de tal o cual DMU son más importantes o menos costosos que los de otra"
Algunas DMUs obtendrán eficiencias relativas del 100% las cuales se denominan DMU de frontera. Es decir son las mejores en la comparación relativa.
El Análisis Envolvente de Datos fija en la frontera varias DMUs. Para evitar que se ubiquen en la frontera demasiadas, se recomienda usar 3 veces más unidades de decisión que variables Input y Output.

3.2 MODELO BCC (BANKERS, CHARNES Y COOPER)

El modelo básico CCR considera que existen rendimientos constantes a escala, permitiendo a las empresas más eficientes ser la referencia de otras empresas con características muy diferentes respecto a la escala de producción. Sin embargo el supuesto de rendimientos constantes a escala no siempre se cumple, es decir puede suceder que algunas DMUs no operen a una escala óptima por la existencia de competencia imperfecta, restricciones financieras, normativas, etc. Para solucionarlos, Bankers, Charnes y Cooper, formularon un modelo que tuviera en cuenta los rendimientos a escala variables y poder así calcular la eficiencia técnica pura (ETP), separándola de los efectos de escala o eficiencia de escala (EE), derivados de utilizar el modelo CRS en las condiciones anteriores. Para imponer la condición de que la comparación se efectúe entre empresas de las mismas características, es necesario incluir una restricción adicional, de convexidad en el modelo. Es decir para obtener un modelo VRS de cualquier orientación, basta agregar una restricción adicional a las especificaciones anteriores:

$\sum_{r=1}^{n} \lambda_r = e\lambda = 1$

Donde e es un vector fila con todos sus elementos igual a 1. Esta restricción asegura que una unidad ineficiente solo sea comparada con unidades productivas de similar tamaño. Sin esta restricción la unidad bajo análisis puede ser comparada con otras sustancialmente mayores o menores. Es decir en el modelo BCC las DMUs ineficientes se comparan únicamente con las unidades eficientes que operan en una escala semejante, Por esta razón también aparecerán más DMUs en la frontera eficiente al emplear el modelo BCC y se construye una frontera más flexible adaptada a las distintas escalas de producción de cada DMU, que identifica su ineficiencia técnica pura, separando esta del efecto de escala. El modelo BCC orientado a outputs, Cooper, Seiford y Tone lo definen de la siguiente forma:

$\max \eta\beta$
Sujeto a:
$X\lambda \leq x_0$
$\eta\beta y_0 - Y\lambda \leq 0$
$e\lambda = 1$
$\lambda \geq 0$

4. DESCRIPCIÓN DE LOS MODELOS DE DISCRIMINACIÓN A COMPARAR.

En esta sección se realiza una descripción un poco más detallada sobre cada uno de los modelos a comparar dentro de este trabajo. En la sección 4.1 se describe un método clásico de discriminación, siguiendo a [3], [5], [6], [10], [11] y [27], en la sección 4.2 el modelo de clasificación con solución aproximada por técnica de optimización propuesto por Almeida en [4] y en la sección 4.3 se aborda el enfoque DEA - DA, propuesto por Sueyoshi en [16] y [17].

4.1 MÉTODO CLÁSICO: REGLA DE DISCRIMINACIÓN PARA DOS GRUPOS VÍA MÁXIMA VEROSIMILITUD

El siguiente resultado debido a Welch (1939), citado por Rencher (1998), a partir del cual se obtienen algunas reglas de clasificación o discriminación.

Sean $f(X/G_1)$ la función de densidad para X en G_1 y (X/G_2) la función de densidad para X en G_2, con G_1 y G_2 las dos poblaciones, sean p_1 y p_2 las probabilidades a priori, donde $p_1 + p_2 = 1$, entonces la regla de discriminación óptima, es decir, la regla que minimiza la probabilidad total de clasificación incorrecta es:

- Asignar la observación X a G_1 si $p_1 f(X/G_1) > p_2 f(X/G_2)$,
- O asignar a G_2, en otro caso.

Esta situación en la práctica es muy poco frecuente, supóngase que se conocen las distribuciones de las dos poblaciones. Sean $f_1(X)$ y $f_2(X)$ las fdp de cada una de las poblaciones, con X vector de observaciones de tamaño $(p \times 1)$. La regla de discriminación máximo verosimilitud para localizar el caso caracterizado por X en alguna de dos poblaciones, consiste en ubicarlo en la población para la cual X maximiza la verosimilitud o probabilidad.

En símbolos, si G_1 y G_2, son las dos poblaciones, entonces se localiza a X en G_i si

$$L_i(X) = \max_j \{L_j\}, \text{ con } i, j = 1,2 \tag{4.1}$$

Esta regla es extendible a cualquier número de poblaciones. En caso de empate, X se asigna a cualquiera de las poblaciones.

4.1.1 Clasificación en poblaciones con matrices de covarianzas iguales.

Supóngase que las poblaciones G_i se distribuyen $N(\mu_i, \Sigma)$, con $i = 1,2$, de manera que la verosimilitud de la i −ésima población es

$$L_i(X) = |2\pi\Sigma|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(X-\mu_i)'\Sigma^{-1}(X-\mu_i)\right\} \tag{4.2}$$

Maximizar la ecuación (4.2) equivale a obtener el mínimo de $(X-\mu_i)'\Sigma^{-1}(X-\mu_i)$, el cual es la distancia de Mahalanobis de X a μ_i. Se asigna el individuo representado por X, a la población más cercana en términos de ésta distancia; es decir, se asigna el caso X al grupo G_1 si

$$(X-\mu_1)'\Sigma^{-1}(X-\mu_1) \leq (X-\mu_2)'\Sigma^{-1}(X-\mu_2) \tag{4.3}$$

O al G_2si

$$(X-\mu_1)'\Sigma^{-1}(X-\mu_1) > (X-\mu_2)'\Sigma^{-1}(X-\mu_2) \tag{4.4}$$

Al desarrollar (4.3) y simplificar algunos términos, se obtiene que se asigna X a G_1 si

$$(\mu_1-\mu_2)'\Sigma^{-1}X-\frac{1}{2}(\mu_1+\mu_2)'\Sigma^{-1}(\mu_1-\mu_2) > 0 \tag{4.5}$$

O a G_2 en caso contrario. El primer término de (4.5) es la función discriminante lineal, si se llama $b=\Sigma^{-1}(\mu_1-\mu_2)$, entonces la función discriminante es de la forma $Y=b'X$, la cual es una combinación lineal de las medidas asociadas con las variables para un objeto o individuo particular.

Las reglas de ubicación son entonces:

- Si $b'(X-\mu_c) \geq 0$, entonces X se asigna a G_1; o
- Si $b'(X-\mu_c) < 0$, entonces X se asigna a G_2 (4.6)

Donde $\mu_c=\frac{1}{2}(\mu_1+\mu_2)$

Observación: La combinación lineal contenida en $b'X$, fue sugerida por R.A. Fisher (1936), de tal forma que la razón de las diferencias en las medias de las combinaciones lineales a su varianza sea mínima. Esto es, la combinación lineal es la forma $Y=b'X$ y se requiere encontrar el vector de ponderaciones de b que maximice la separación entre los dos grupos

$$\frac{(b'\mu_1-b\prime\mu_2)^2}{b'\Sigma b} \tag{4.7}$$

Manteniendo constante la varianza de la combinación lineal $b'X$; es decir, $var(b'X)=b'\Sigma b$ por multiplicadores de lagrange se concluye que b es proporcional a $\Sigma^{-1}(\mu_1-\mu_2)$.
Hasta ahora se ha asumido que las dos poblaciones se conocen a través de su distribución, en la práctica los parámetros que las determinan e identifican se estiman e infieren desde muestras aleatorias independientes.

Supóngase que se extraen las muestras $X_{1(i)}, \dots, X_{ni(i)}$ de una población $N(\mu_i, \Sigma)$ para $i = 1,2$. Con base en esta información se pretende asignar la observación X a G_1 o G_2. Los estimadores para μ_i y Σ respectivamente son:

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{j(i)}/n_i, \;\; i = 1,2$$
$$s = \frac{1}{n_1+n_2-2}\left[\sum_{j=1}^{n_1}\left(X_{j(1)} - \bar{X}_{(1)}\right)\left(X_{j(1)} - \bar{X}_{(1)}\right)' + \sum_{j=1}^{n_2}\left(X_{j(2)} - \bar{X}_2\right)\left(X_{j(2)} - \bar{X}_2\right)'\right] \qquad (4.8)$$

Al sustituir estas estimaciones en (4.5), la función discriminante muestral toma la forma $\hat{Y} = \hat{b}'X$, se usan los mismos criterios dados en (4.6), con los datos muestrales los criterios son:

Si $\hat{b}'X \geq \hat{b}'\bar{X}_c$, X se asigna a G_1 o,
Si $\hat{b}'X < \hat{b}'\bar{X}_c$, X se asigna a G_2, (4.9)

Con $\bar{X}_c = \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$ y $\hat{b} = S^{-1}(\bar{X}_1 - \bar{X}_2)$

4.1.2 Clasificación en poblaciones con matrices de covarianzas distintas.

Si las dos poblaciones G_1 y G_2 tienen distribución normal p variante con matrices de covarianzas distintas $\sum_1 \neq \sum_2$, el logaritmo de la razón de la verosimilitud para una observación particular X es el siguiente.

$Q(X) = \beta + \gamma X + X\Lambda X$ Con

$$\beta = \frac{1}{2} ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}\left(\mu'_1 \Sigma_1^{-1} \mu_1 - \mu'_2 \Sigma_2^{-1} \mu_2\right) \qquad (4.10)$$

$$\gamma = \left(\mu'_1 \Sigma_1^{-1} - \mu'_2 \Sigma_2^{-1}\right)$$

$$\Gamma = \left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)$$

En la expresión $Q(X)$ el último término $X'\left(\sum_1^{-1} - \sum_2^{-1}\right)X$, corresponde a los cuadrados y productos cruzados de las componentes del vector $X, Q(X)$ se denomina función de discriminación cuadrática. Nótese que si $\sum_1 = \sum_2$ entonces $Q(X)$ con la función discriminante lineal.

El criterio para clasificar una observación X es el siguiente:

Si $Q(X) \geq 0$, entonces X se asigna a G_1; o

Si $Q(X) < 0$, entonces X se asigna a G_2.

En términos muestrales, si se obtiene una muestra de la población G_1 y una de la población G_2, se calcula un valor muestral de $\hat{Q}(X)$ al reemplazar μ_i por $\bar{X}_i$ y $\sum_i$ por $\boldsymbol{S_i}$ (i-ésimo grupo), ésta es:

$$\hat{Q}(X) = b + c'X - X'AX \tag{4.11}$$

La regla para clasificar una observación muestral X es similar al caso poblacional como se indica en el recuadro anterior; es decir, se asigna la observación o individuo X al grupo G_1 si $\hat{Q}(X) \geq 0$ y al grupo G_2 en caso contrario.

Cuando $\sum_1 \neq \sum_2$ la función de clasificación cuadrática $\hat{Q}(X)$ es óptima de manera asintótica; aunque para muestras de tamaño pequeño $\boldsymbol{S_i}$ no es un estimador estable de $\sum_i$, es decir $\boldsymbol{S_i}$ varía bastante en muestras de la misma población o grupo. En tales casos Rencher (1998, pág. 233) recomienda emplear la regla de discriminación lineal. Para muestras de tamaño grande y con amplias diferencias en las matrices de varianza, la función de discriminación cuadrática es la más recomendable.

4.2 MODELO DE CLASIFICACIÓN CON SOLUCIÓN APROXIMADA POR TÉCNICA DE OPTIMIZACIÓN.

El modelo propuesto por Almeida en [4] parte del hecho de que el grado de inserción que posiciona una DMU es medido sobre una Función de Distribución o Densidad de Probabilidad (fdp), cuando ésta es conocida. Pero cuando no se conocen las fdp se pueden obtener diversas soluciones aproximadas. Se propone una aproximación a través de la función que define las curvas de indiferencia. De esta manera se construye la función *h*:

$$\frac{X_{iI} - X_{iE}}{X_{iI}} = h, \tag{4.12}$$

Para el caso de dos dimensiones y para más de dos dimensiones es:

$$\frac{X_{mI} - X_{mE}}{X_{mI}} = h \tag{4.13}$$

Donde h se usa para comparar DMUS de un mismo grupo. Se tiene entonces una escala creciente partiendo de cero, indicando cuales DMUS se encuentran más hacia adentro.
Para establecer una medida relativa de posicionamiento de las curvas de indiferencia, incluyendo las características de dispersión para cada población definiendo así el procedimiento para la clasificación de DMUS en diferentes grupos se tiene la siguiente ecuación:

$$h_{Go} = \frac{(1 - h_m)h_0}{(1 - h_0)h_m} \tag{4.14}$$

Aquí h_{Go} Será nuestro puntaje en el proceso de comparar DMUS y se encuentra en función de h_0 y h_m, donde h_m es un valor calculado para la DMU P_m que se encuentra en el centro de la distribución y h_0 es una valor calculado para la DMU P_O en análisis. h_{Go}

Identifica la curva de indiferencia de P_O tanto en relación al extremo como al centro de la distribución. Ver gráfico 2.
No hemos calculado el valor de probabilidad de la DMU, pero si hemos determinado su posición.
Resulta un valor h_G para cada grupo posicionando a la DMU en una diferente curva de indiferencia, conforme al grupo.
Una DMU será clasificada en el grupo que presente mayor valor de h_{Go}.

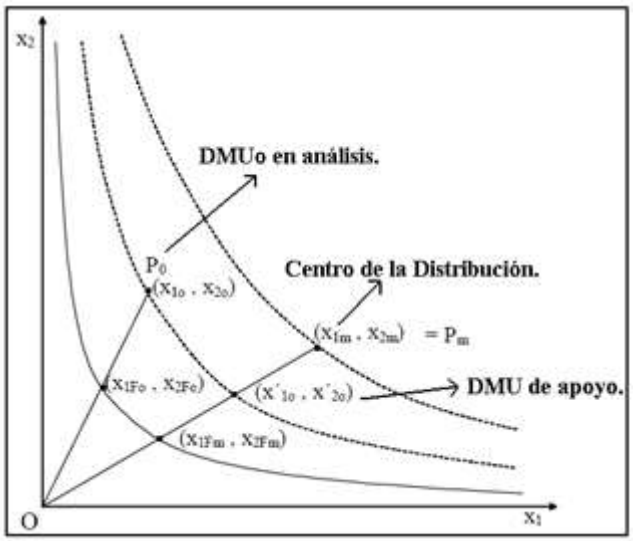

Gráfico 2: Posicionamiento de las curvas de indiferencia

Se pretende obtener el valor de *h* a través de los polígonos cuyos vértices están sobre las curvas de indiferencia.

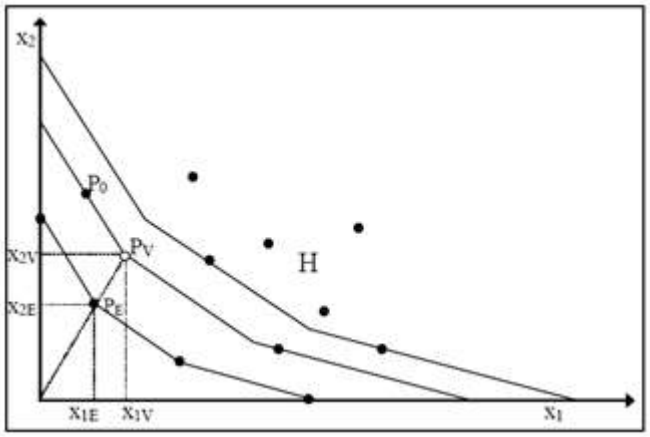

Gráfico 3: Polígonos de Indiferencia

El gráfico 3, es una representación bidimensional. La envolvente obtenida a partir de la muestra está apoyada en cuatros DMUS, donde identificamos el punto P_E .
Cuando construimos los polígonos, ellos son representativos de los polígonos de indiferencia, el procedimiento se origina de una medida DEA denominada Distancia Direccional**.** Esta medida es la que también denominamos ***h*** y se obtiene de la siguiente ecuación.

$$h = \frac{X_{1V} - X_{1E}}{X_{1V}} = \frac{X_{2V} - X_{2E}}{X_{2V}} \qquad (4.15)$$

DEA me da el valor de esta medida. Cuando *h=0* son DMUS situadas sobre la envolvente y cuando *0<h<1* son DMUS internas.

De esta manera el modelo es el siguiente:
max h

$Sujeto\ a$:

$$\sum_{j=1}^{n} \lambda_j X_{ij} + hX_{ijo} \leq X_{ijo} \qquad (4.16)$$
$$\sum_{j=1}^{n} \lambda_j = 1$$
$$\lambda_j \geq 0$$

La ecuación (4.17) explicita la solución de (4.16):

$$h = \frac{(X_{ijo} - X_{ijf})}{X_{ijf}} \qquad i = 1,2,\dots,m \qquad (4.17)$$

El procedimiento para definir la medida de posición de las DMUS de las muestras de calibración de cada grupo relativamente a su frontera se describe en los siguientes ocho pasos:

1. **Efectuar el ajuste de las variables originales:**
   En el modelo propuesto se considera que las variables de clasificación son limitadas, inferior o superiormente. Esto se puede traducir en el hecho de que una población puede estar influenciada por dos tipos de variables, una de características favorables (VF) y otra de características desfavorables (VD). De esta hipótesis resulta una frontera, la cual divide la región en dos partes: una es la región H la cual define el dominio de las variables y la otra es la región I donde las condiciones son extremadamente desfavorables a la ocurrencia de una entidad de esta población.
   Para estimar el límite superior de una variable se utiliza el estimador de máxima verosimilitud donde n es el tamaño de la muestra: $\hat{Y} = \max(Y_1, Y_2, \dots, Y_n)$ y el estimador de máxima verosimilitud para el límite inferior de una variable es: $\widehat{X} = \min(X_1, X_2, \dots, X_n)$.

De esta manera se substituye cada valor X_{ij} de X_i (variable favorable) por $X_{ij} - \widehat{X_i}$ y cada valor Y_{rj} de Y_r (variable desfavorable) por: $\widehat{Y_r} - Y_{rj}$.
Este proceso se realiza con el fin de trabajar con variables todas con características favorables, las cuales serán definidas por:
$X_i = \{X_i \in R / X_i \geq 0, i = 1,2, \dots, m+s\}$ Delimitando la región.

2. Llamaremos PPL al sistema de ecuaciones generada por (4.16) el cual se ejecutará para cada población en estudio.

3. DMUS de una determinada población que presenten h=0, definirán su frontera. El PPL de ésta población será reformulado, a partir de este momento y solo contendrá las DMUS de la frontera y una particular DMU que será probada o testeada. Ya que la contribución de las otras DMUS deja de ser relevante en la determinación de la medida de h de la DMU a ser testeada. El nuevo sistema será denominado PPLe para esta población (PPL definido sobre la envolvente de esta población) y corresponde al sistema de ecuaciones generada por (4.18).

4. Se obtienen tantos PPL como poblaciones en estudio.

5. Para probar una nueva entidad frente a una cierta población, ella se hará a través del PPLe de esta.

6. El PPLe para cada población, contiene m restricciones, formadas por las m variables y tendrá la siguiente forma al testar una nueva DMU.

PPLe
El índice cero será asignado para la DMU a ser testeada.
El índice j se refiere a las entidades de la envolvente de la particular población a probar.

$$\max\ h$$
$$Sujeto\ a:$$

$$\sum_{j \in envolvente}^{n} \lambda_j X_{ij} + hX_{ijo} \leq X_{ijo} \quad i = 1,2,..,m \ \ (m\ lineas)$$
$$\sum_{j \in envolvente}^{n} \lambda_j = 1$$
$$\lambda_j \geq 0 \qquad (4.18)$$

- Para probar una nueva unidad se adoptará el siguiente sistema de comparación.
- La DMU es introducida apenas como un dato adicional X_{ijo} en las inecuaciones del PPLe de la población en la que ella será probada, después del ajuste de las variables.
- La DMU no debe ser introducida como una restricción adicional en $\sum_{j \in envolvente}^{n} \lambda_j X_{ij}$ de este modo queda imposible la inclusión de ésta DMU como una entidad de la envolvente, impidiendo la modificación en la forma actual de la envolvente (índice $j \neq 0$).

7. **Definición de h:**

- La DMU se encuentra sobre la envolvente actual o en el interior de la región de dominio H, si resulta PPLe viable con $h \geq 0$.
- Una nueva DMU está fuera de la región H, si resulta PPLe viable con $h < 0$.

8. **Clasificación de nuevas entidades:**
   Una vez definidas las fronteras que delimitan las poblaciones en estudio, las DMUS actuales y nuevas de cualquiera de las poblaciones podrá ser testeada para cualquiera de los siguientes dos casos:

- Verificación de la posibilidad de una región común a dos o más poblaciones. Así DMUS que están dentro de la región común, presentarán un $h_k \geq 0$, medidas en relación a cada una de las k fronteras de las k poblaciones. Las DMUS que están fuera de una de las fronteras presentarán $h_k < 0$ en relación a esta frontera.
- Clasificación de nueva DMU a través de los respectivos PPLe de cada población. Determinando en cual población esta DMU debe ser clasificada. En el caso de que esta misma no esté localizada en la región común. La clasificación obedece al criterio de mayor h_G , es decir la nueva DMU será clasificada en el grupo donde el PPLe resulte mayor h_G.

4.3 MODELO DEA-DA EXTENDIDO

El tercer modelo a comparar en este estudio es llamado "DEA-Análisis Discriminante Extendido" también referido en la literatura como modelo de programación lineal (LP) el cual es una modificación de la formulación original DEA-DA y fueron planteados por Sueyoshi. Las características más importantes de estos dos modelos, desarrollados ampliamente en [16] y [17], se expresan a continuación:

4.3.1 DEA-Análisis Discriminante (DEA-DA) desde la visión de la Programación por metas.

Según [16] Esta formulación incorpora dentro del marco del Análisis Discriminante (DA) una fuerte metodología del DEA. Este modelo puede ser definido en las dos etapas siguientes:

a. Clasificación e identificación del traslapo (COI)
b. Manejo del traslapo (HO)

a. Clasificación y manejo del traslapo, primera etapa (COI):

$$min\ \varphi = \sum_{j \in G_1} S_{Ij}^{+} + \sum_{j \in G_2} S_{2j}^{-}$$

$$s.t\ \sum_{i=1}^{k} \alpha_i z_{ij} + S_{Ij}^{+} - S_{Ij}^{-} = d, \qquad j \in G_1$$

$$\sum_{i=1}^{k} \beta_i z_{ij} + S_{2j}^{+} - S_{2j}^{-} = d - \eta, \quad j \in G_2$$

$$\sum_{i=1}^{k} \alpha_i = 1,$$

$\sum_{i=1}^{k} \beta_i = 1$

Todas las holguras ≥ 0, $\alpha_i \geq 0, \beta_i \geq 0$, d sin restricción y η un número positivo muy pequeño.

Una nueva muestra es denotada por z_{im} y puede ser identificada en el traslapo si cumple con el siguiente criterio:

$Si \ \sum_{i=1}^{k} \lambda_i^* z_{im} > d^* \geq \sum_{i=1}^{k} \beta_i^* z_{im}$

O $\sum_{i=1}^{k} \lambda_i^* z_{im} \leq d^* < \sum_{i=1}^{k} \beta_i^* z_{im}$

Una nueva muestra z_{im}, puede ser identificada en el G_1 si cumple con el siguiente criterio:

$\sum_{i=1}^{k} \lambda_i^* z_{im} \geq d^*$ y $\sum_{i=1}^{k} \beta_i^* z_{im} \geq d^*$

Una nueva muestra z_{im}, puede ser identificada en el G_2 si cumple con el siguiente criterio:

$\sum_{i=1}^{k} \lambda_i^* z_{im} < d^*$ y $\sum_{i=1}^{k} \beta_i^* z_{im} < d^*$

b. Manejo del traslapo (HO)

$min \ \varphi = \sum_{j \in G_1} S_{Ij}^{+} + \sum_{j \in G_2} S_{2j}^{-}$

$s.t \ \sum_{i=1}^{k} \alpha_i z_{ij} + S_{Ij}^{+} - S_{Ij}^{-} = d, \qquad j \in G_1$

$\sum_{i=1}^{k} \alpha_i z_{ij} + S_{2j}^{+} - S_{2j}^{-} = d - \eta, \ \ j \in G_2$

$\sum_{i=1}^{k} \alpha_i = 1$

Todas las holguras ≥ 0, $\alpha_i \geq 0$, d sin restricción.

$si \ \sum_{i=1}^{k} \lambda_i^* z_{im} \geq d^*, \ \text{j} \in G_1 \cap G_2, \ \text{entonces} \ j \in G_1$

$si \ \sum_{i=1}^{k} \lambda_i^* z_{im} < d^*, \ \text{j} \in G_1 \cap G_2, \ \text{entonces} \ j \in G_2$

4.3.2 Modelo DEA-DA Extendido.

Según [17] Matemáticamente el enfoque DEA-DA Extendido tiene el siguiente proceso de computación en dos etapas.

a. Clasificación e identificación del traslapo (COI)
b. Manejo del traslapo (HO)

a. Clasificación y manejo del traslapo, primera etapa (COI):

$Minimizar \ \ \sum_{j \in G_1} s_{1j}^{+} + \sum_{j \in G_2} s_{2j}^{-}$

$Sujeto\ a \quad \sum_{i=1}^{k}(\lambda_i^+ - \lambda_i^-)z_{ij} + s_{ij}^+ - s_{1j}^- = d + 1, \quad j \in G_1$

$\sum_{i=1}^{k}(\lambda_i^+ - \lambda_i^-)z_{ij} + s_{2j}^+ - s_{2j}^- = d, \ j \in G_2$

$\sum_{i=1}^{k}(\lambda_i^+ - \lambda_i^-) = 1,$

$slacks \geq 0, \lambda_i^+ \geq 0, \lambda_i^- \geq 0, d \sin restricciones$

Donde:

$s_{1j}^+ \ y \ s_{1j}^- \ \ (j \in G_1)$, Son desviaciones positivas y negativas de la función discriminante lineal de un puntaje discriminante (d) del primer grupo.

s_{2j}^+ y $s_{2j}^- \ (j \in G_2)$, son desviaciones positivas del segundo grupo.

$s_{1j}^+ \ y \ s_{2j}^-$, indica una ocurrencia de clasificación incorrecta, mientras el restante de slacks denota correcta clasificación.

Todos los factores observados z_{ij}, están conectados por $\sum_{i=1}^{k}\lambda_i z_{ij}$, donde λ_i, es un peso del i-ésimo factor.

Estos pesos están restringidos en la forma que la suma de los valores absolutos de $\lambda_i (= \lambda_i^+ - \lambda_i^-)$, es la unidad.

$\lambda_i^* (= \lambda_i^{+*} - \lambda_i^{-*}) \ y \ d^*$, son las soluciones óptimas del modelo anterior. Entonces la nueva r-ésima observación muestreada $Z_r = (z_{1r}, \dots, z_{kr})^T$, está clasificada por la siguiente regla:

$Si \ \sum_{i=1}^{k}\lambda_i^* z_{ir} \geq d^* + 1, \qquad entonces\ la\ r - ésima\ observación \in G_1$

$Si \ d^* + 1 > \sum_{i=1}^{k}\lambda_i^* z_{ir} > d^*, \quad entonces\ la\ r - ésima\ observación \in G_1 \cap G_2$

$Si\ d^* \geq \sum_{i=1}^{k}\lambda_i^* z_{ir} \qquad entonces\ la\ r - ésima\ observación \in G_2$

Basándonos en las tres reglas anteriores todas las observaciones quedan clasificadas en los siguientes subconjuntos:

$C_1 = \{j \in G_1 | \sum_{i=1}^{k}\lambda_i^* z_{ij} \geq d^* + 1\},$

$C_2 = \{j \in G_2 | \sum_{i=1}^{k}\lambda_i^* z_{ij} \leq d^*\}$

$D_1 = G_1 - C_1 \ \ y \ \ D_2 = G_2 - C_2$

b. Tratamiento del traslapo, segunda etapa (HO):

Un nuevo puntaje discriminante c es incorporado dentro del modelo anterior que es usado solo para reclasificar los dos subgrupos $(D_1 \cup D_2)$, matemáticamente, el proceso HO es formulado como sigue:

$$Minimizar \quad \sum_{j\in D_1} s_{1j}^{+} + \sum_{j\in D_2} s_{2j}^{-}$$

$$Sujeto\ a \quad \sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})z_{ij} \geq d+1, \quad j \in C_1$$

$$\sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})z_{ij} + s_{ij}^{+} - s_{1j}^{-} = c, \quad j \in D_1$$

$$\sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})z_{ij} + s_{2j}^{+} - s_{2j}^{-} = c, \ \ j \in D_2$$

$$\sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})\, z_{ij} \leq d, \ j \in C_2$$

$$\sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-}) = 1,$$

$$d \leq c \leq d+1$$

$$slacks \ \geq 0, \lambda_i^{+} \geq 0, \lambda_i^{-} \geq 0, d\ y\ c\ \ \text{sin}\, restricciones$$

Toda observación clasificada correctamente están restringidas por

$$\sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})z_{ij} \geq d+1, j \in C_1 \text{ y } \sum_{i=1}^{k}(\lambda_i^{+} - \lambda_i^{-})\, z_{ij} \leq d, \ j \in C_2$$

Todos los slacks con respecto a estas restricciones son omitidos de la función objetivo.

Bajo la restricción entre $d+1$ (puntaje discriminante para $c1$) y d (puntaje discriminante para $c2$), el nuevo puntaje discriminante c es determinado para minimizar la desviación total de las observaciones en el traslapo.

Dado c^* y $\lambda_i^*(=\lambda_i^{+*} - \lambda_i^{-*})$, siendo el puntaje discriminativo y el peso estimado obtenido de la optimalidad del modelo anterior, respectivamente.

La nueva observación muestreada, identificada como un miembro del traslapo en el paso previo, puede ser clasificada por la siguiente regla:

$$Si \ \ \sum_{i=1}^{k}\lambda_i^* z_{ir} \geq c^*, \qquad entonces\ la\ \ observación \in G_1$$

$$Si \ \ \sum_{i=1}^{k}\lambda_i^* z_{ir} < c^*, \quad entonces\ la\ \ observación \in G_2$$

Para un caso en donde la igualdad es observada en la ecuación, dependemos sobre una segunda opinión, con respecto a su clasificación.

5. CONJUNTO DE DATOS

Para comparar el desempeño de los tres modelos discriminantes, se usaron en este trabajo dos conjuntos de datos: El primer conjunto se encuentra desenvuelto en [4] y hace referencia a la manera en como las empresas clientes realizan compras, este primer conjunto de datos es usado en [6]. El segundo conjunto de datos hace referencia a la información de bancos japoneses y fueron usados en [16] y [17].

a. Conjunto 1:

En [4], se presenta un particular problema que envuelve dos grupos y está constituido de 14 variables. Se buscó discriminar la variable X_{11} a partir de las variables discriminatorias X_1, X_3 y X_7, identificadas como las más adecuadas para este fin.
Se procura interpretar la manera en como las **empresas clientes** efectúan sus compras, si basadas en las características específicas del producto o si emplean el análisis del valor total de la compra. Así la **empresa distribuidora** puede alterar presentaciones de ventas y beneficios ofrecidos, conforme la característica de cada empresa compradora, mejorando el desempeño de las ventas, así:

X_1 = Velocidad de entrega de los pedidos.
X_3 = Flexibidad de precios, según los representantes dispongan una política de precios de la empresa.
X_7 = Calidad del producto.
X_{11}= Variable con dos categorías. X_{11}=0, si el comprador se basa apenas en las características específicas del producto (identificando como el comprador del grupo 0). $X_{11} = 1$, si emplea un análisis del valor total de la compra (identificando como el comprador del grupo 1).
La muestra se encuentra constituida de 100 observaciones, un total de 40 compradores del grupo 1 y 60 compradores del grupo 2. El conjunto de datos es mostrado en la tabla 2.

Grupo 0				Grupo 1			
X1	X3	X7	X11	X1	X3	X7	X11
1,8	6,3	8,4	0	5,5	9,4	7,6	1
1,9	7,9	9,7	0	3,4	9,7	4,8	1
1,3	6,2	6,9	0	4,9	7,7	5,9	1
2,8	8,1	6,6	0	5,3	9,7	6,8	1
2,4	6,7	7,2	0	3,3	8,6	6,3	1
3	6	8	0	3,4	8,3	5,2	1
1,8	7,5	7,6	0	4	9,1	7,3	1
0	6,9	8,9	0	5	8,6	3,7	1
2	6,5	8,5	0	5	9,4	6,3	1
3,4	5,6	9,1	0	3,8	8,7	5,6	1
2,6	8,2	9	0	2,9	7,7	7,7	1
4,5	6,3	8,8	0	5,1	9,2	4,5	1
2,8	6,7	9,2	0	4,1	9,3	7,4	1
1,1	7,2	10	0	3	5,5	6	1
1,6	6,4	7,1	0	3,7	9	6,8	1
2,3	8,3	9,1	0	5,3	8,5	4,8	1
3,6	5,9	8,4	0	5,2	9,1	7,3	1
1	7,1	9,9	0	4,2	9,4	8,5	1
2,5	7	9	0	3,8	8,3	5,3	1
2,9	7,3	8	0	3,3	9,7	5,2	1
0,6	6,4	8,4	0	3	7,8	7,9	1
3,1	6,7	8,4	0	2,5	9	6	1
3,4	5,7	8,2	0	4,1	6,9	5,2	1
2,7	7,1	7,8	0	6	9,6	4,5	1
4	6,5	8,7	0	4,6	9,5	7,6	1
2,4	6,6	7,2	0	2,4	8,8	5,8	1
4,1	5,9	8,4	0	3,9	9,1	8,3	1
3,4	6,4	8,4	0	3,7	8,6	6,7	1
2,4	7,7	6,2	0	4,7	9,9	6,8	1
3,6	5,8	9,3	0	3,2	5,7	6,2	1
2,4	6,4	8,8	0	4,7	9,9	6,8	1
1,9	7,6	7,7	0	3	9,1	8,4	1
4,9	7,4	9,6	0	5,1	8,7	3,8	1
2,3	8	8,7	0	4,6	7,9	4,7	1
3	6,6	8,2	0	5,2	9,7	6,7	1
1,6	6,1	8,2	0	3,5	9,9	5,4	1
2,6	8,5	6,8	0	2,8	8,9	8,2	1
2,4	8,4	6,7	0	5,2	9,3	4,6	1
1,9	5	8,2	0	5,9	9,6	4,5	1
2	5,2	8,4	0	4,9	9,3	6,2	1
				3,1	10	3,8	1
				5,8	8,8	6,7	1
				5,4	8	5,2	1
				3,7	8,2	5,2	1
				5,4	9,6	7,7	1
				4,3	7,6	4,4	1
				3,1	9,9	3,8	1
				4,2	9,2	7,3	1
				5,6	8,2	5,3	1
				3,6	9,9	4,9	1
				4,5	8,7	6,8	1
				5,5	8,7	4,9	1
				3,4	5,5	6,3	1
				2,3	7,6	7,4	1
				2,1	7,4	7,2	1
				4,3	9,3	7,4	1
				4,8	7,6	5,8	1

Tabla 2: Conjunto 1 de datos de Empresas Clientes al momento de efectuar sus compras.

a. Conjunto 2:

En [17], es usado un conjunto de datos reales relacionado con el rendimiento de bancos japoneses y que contiene 100 observaciones, los datos fueron tomados de una revista de negocios muy reconocida en Japón llamada "Financial Business (in Japanese)" (Septiembre 1997, pp. 44-47) cuyos suscriptores son usualmente líderes corporativos. Todas las observaciones en el conjunto de datos son listadas por sus rendimientos corporativos. Para ello, G1 contiene 50 datos, los cuales son los mejores 50 bancos clasificados, mientras que el grupo G2 contiene los 50 bancos restantes mejores clasificados. Se midió 7 índices de rendimiento. El conjunto de datos es mostrado en la tabla 3.

$X_1 =$ Retorno de los activos totales (=total de las ganancias/promedio de los activos totales),
$X_2 =$ Equidad de activos totales (=equidad total/promedio de los activos totales),
$X_3 =$ Relación Costo beneficio (=total de gastos de funcionamiento/total de ganancias),
$X_4 =$ Retorno sobre los activos nacionales total (=total ganancias nacionales/Promedio total de activos nacionales)
$X_5 =$ Tasa de mal préstamo (total mal prestamos/total prestamos)
$X_6 =$ Tasa de pérdida de mal crédito (malos préstamos despreciados como pérdidas/total de malos préstamos)
$X_7 =$ Retorno sobre equidad (ingresos disponibles para común/promedio equidad)

Financial Performance of Japanese Banks							
Return On total assets x1	**Equity total assets x2**	**Cost-Profit Rate x3**	**Return on total domestic assets x4**	**Bad loan Ratio x5**	**Loss ratio Of bad loan x6**	**Return on equity x7**	**Grupo**
0,67	9,28	48,28	0,36	2,08	46,99	18,68	**1**
1	4,78	61,98	0,94	0,31	155,2	27,26	**1**
0,65	9,1	50,07	0,52	2,41	48,35	20,9	**1**
1,67	4,21	51,69	0,88	2,97	80,1	51,24	**1**
0,91	10,64	59,72	0,72	1,08	78,64	21,44	**1**
0,62	8,75	51,9	0,35	2,49	46,18	19,83	**1**
0,71	8,75	53,75	0,52	3,64	51,52	24,87	**1**
1,06	6,58	59,31	0,93	1,37	47,59	19,56	**1**
1,02	10,15	58,59	0,81	1,34	65,4	25,79	**1**
0,87	9,16	61,21	0,84	1,55	77,93	21,45	**1**
1,13	6,23	58,33	0,95	3,48	37,44	23,78	**1**
0,82	11,28	61	0,69	1,52	45,85	17,1	**1**
0,64	9,22	53,15	0,19	4,64	55,55	19,93	**1**
0,88	11,14	61,84	0,63	1,7	52,2	14,19	**1**
0,77	9,54	58,62	0,63	0,61	49,62	15,15	**1**
0,95	9,45	58,77	0,84	3,48	19,53	14,96	**1**
0,73	9,24	59,93	0,59	1,26	61,95	15,56	**1**

0,8	10,04	62,9	0,6	2,18	65,03	19,41	**1**
0,83	9,6	57,83	0,63	2,66	47,66	19,46	**1**
0,53	8,92	64,13	0,38	3,88	50,26	18,16	**1**
1,05	6,38	66,56	0,98	3,7	48,67	20,39	**1**
0,91	10,17	60,03	0,75	2,13	36,63	19,94	**1**
1,05	8,44	57,93	0,67	2,48	48,67	30,51	**1**
0,8	9,86	67,12	0,68	1,33	69,61	19,04	**1**
1,05	4,53	59,6	0,79	3,09	50,72	29,15	**1**
0,9	9,77	61,37	0,71	2,26	62,25	110	**1**
0,95	4,41	63,71	0,84	2,8	56,11	35	**1**
0,69	9,64	70,12	0,49	0,64	85,12	12,77	**1**
0,78	11,83	67,01	0,49	0,36	58,41	19,4	**1**
0,56	13,61	66,89	0,34	0,81	51,89	9,43	**1**
0,79	9,35	68,02	0,66	0,79	59,32	18,46	**1**
0,73	9,18	66,07	0,58	1,15	66,93	18,82	**1**
0,86	9,09	61,1	0,6	1,95	47,28	16,97	**1**
1	11,33	64,34	0,46	1,19	46,54	18,44	**1**
0,96	5,06	63,42	0,88	4,66	37,96	26,09	**1**
0,76	9,19	59,35	0,48	1,8	22,89	15,28	**1**
1,72	4,41	52,91	0,59	6,94	40,62	49,3	**1**
1,09	9,03	63,24	0,63	4,19	42,71	26,44	**1**
0,66	9,06	68,22	0,73	1,5	37,58	16,17	**1**
1,57	4,28	52,91	0,65	6,87	33,21	45,93	**1**
0,68	10,19	68,76	0,52	0,97	57,65	15,21	**1**
0,53	8,7	63,42	0,35	2,9	62,7	18,48	**1**
0,66	9,63	68,17	0,49	1,1	57,2	17,4	**1**
0,93	4,35	65,66	0,81	1,76	39,3	28,58	**1**
1,18	8,69	52,83	0,54	6,8	43,16	35	**1**
0,7	4,72	71,29	0,9	2,39	49,78	18,33	**1**
0,71	9,35	69,69	0,64	1,84	38,16	16,5	**1**
0,65	9,11	65,96	0,56	1,42	44,16	16,72	**1**
0,71	10,54	66,37	0,47	1,43	43,62	14,23	**1**
0,72	9,47	69,35	0,48	0,86	56,36	15,34	**1**
0,71	9,31	66,43	0,51	1,66	52,79	16,04	**2**
0,69	8,42	67,94	0,55	0,38	45,38	18,42	**2**
0,55	9,44	70,29	0,43	1,24	74,95	12,59	**2**
0,51	9,09	61,15	0,24	4,7	68,15	20,74	**2**
0,6	10,69	66,77	0,41	2,42	40,55	12,41	**2**
1,1	3,75	62	0,5	2,21	49,9	39,62	**2**
0,72	4,2	66,99	0,73	1,53	45,88	23	**2**
0,64	9,63	67,67	0,47	2,82	50,61	18,09	**2**
0,62	9,26	68,59	0,49	1,51	45,25	16,72	**2**
0,58	9,82	71,67	0,42	0,89	64,84	14,43	**2**
0,83	4,96	65,84	0,74	3,65	39,98	23,54	**2**
0,66	10,85	71,15	0,34	1,19	62,33	13,44	**2**
0,88	4,28	68,67	0,78	3,93	46,31	26,11	**2**
0,99	4,37	64,49	0,87	5,48	21,93	26,11	**2**
0,72	4,54	72,42	0,59	1,54	71,95	21,39	**2**
0,87	4,25	66,45	0,59	2,18	56,39	28,51	**2**
0,96	4,02	64,27	0,75	3,59	80,51	44,48	**2**
0,67	6,06	70,75	0,39	1,42	69,34	13,82	**2**
0,68	8,76	70,38	0,49	1,34	49,27	16,19	**2**
0,85	4,24	66,33	0,64	3,03	48,97	28,41	**2**
0,6	9,2	71,87	0,44	1,09	59,29	16,26	**2**

0,52	9,08	69,43	0,34	1,52	55,68	14,09	**2**
0,65	9,22	67,79	0,4	1,47	62,74	16,7	**2**
0,76	8,13	67,71	0,47	1,71	47,88	19,42	**2**
0,57	9,63	70,41	0,39	2,72	52,91	12,36	**2**
0,56	9,43	74,88	0,39	0,54	59,3	15,04	**2**
0,59	10,55	71,99	0,35	0,89	35,83	11,86	**2**
0,75	8,42	64,59	0,48	3,1	50,94	20,39	**2**
0,6	9,02	66,75	1,13	4,65	34,23	18,6	**2**
0,8	4,16	69,81	0,69	2,62	33,13	26,04	**2**
0,65	8,29	71,15	0,52	2,23	52,1	16,41	**2**
0,64	4,11	72,33	0,49	0,86	57,3	20,6	**2**
0,49	9,05	74,25	0,43	1,61	51,57	10,16	**2**
1,12	4,24	64,41	0,42	2,71	31,45	32,12	**2**
0,64	9,03	71,71	0,34	1,3	57,4	16,89	**2**
0,45	10,43	79,26	0,4	1,28	65,05	10,72	**2**
0,63	9,7	71,05	0,21	0,68	46,2	13,57	**2**
0,6	8,4	65,9	0,4	3,09	33,48	20,68	**2**
0,87	4,77	64,63	0,44	3,16	35,94	25,03	**2**
0,6	4,83	74,85	0,38	0,27	51,76	15,28	**2**
0,6	4,11	73,99	0,37	1,12	80,03	17,97	**2**
0,73	3,46	71,56	0,48	2,18	67,48	24,89	**2**
0,72	6,88	76,31	0,26	0,56	48,12	14,29	**2**
0,75	4,47	72,36	0,42	2,46	63,08	20,8	**2**
0,83	4,13	60,99	0,37	2,63	32,01	28,67	**2**
0,88	4,13	59,07	0,45	5,88	29	28,49	**2**
0,82	4,65	65,99	0,47	3,26	28,62	23,28	**2**
1,01	4,53	63,39	0,3	4,46	43,69	29,04	**2**
0,4	9,1	74,18	0,33	3,83	54,38	13,98	**2**
0,54	4,47	78,35	0,49	0,96	46,59	14,63	**2**

Tabla 3: Conjunto2 de datos, Rendimientos financieros de 100 bancos japoneses

6. METODOLOGÍA DE COMPARACIÓN

La evaluación de los desempeños de los modelos de discriminación se llevó a cabo utilizando las diferentes Tasas de Clasificación Errónea definidas en el capítulo 2 en la sección 2.2.

Los tres modelos de discriminación se compararon llevando a cabo el siguiente conjunto de pasos:

1. Para cada conjunto de datos se realizó un análisis exploratorio multivariado, evaluando matrices de correlación y diagramas de dispersión. Se realizó adicionalmente una verificación sobre los supuestos de Multinormalidad, homogeneidad de varianzas e igualdad de medias.
2. Cada conjunto de datos fue dividido en dos partes: Del 100% de los datos se tomó de forma aleatoria el 70% para conformar el conjunto de calibración o entrenamiento y así construir las funciones discriminantes de cada uno de los modelos comparados en este trabajo.
3. Con el 30% de los datos restantes se conformó el conjunto de validación, todas las muestras de este conjunto se clasificaron mediante las funciones obtenidas en el paso anterior.
4. Se calculó la tasa de clasificación errónea para el conjunto de validación del paso anterior para cada modelo de discriminación usado dentro de este trabajo.
5. Para el modelo clásico de discriminación se calcularon las funciones lineales y cuadráticas de Fisher y de forma adicional se calcularon las tasa de clasificación errónea por Resustitución, validación cruzada y Bootstrap.
6. Los cálculos computacionales se realizaron con R (R Development Core Team 2018), para las funciones lineales de Fisher se usó la función lda(), para las funciones cuadráticas qda(), para realizar predicciones predict(), para resolver los modelos de programación lineal y obtener los parámetros para las funciones de clasificación se usó la función simplex(). Todos los algoritmos fueron enumerados y se encuentran como anexos en este trabajo.

7. RESULTADOS

En una matriz de correlación, en la medida que los valores se acercan a cero, la correlación es menor y en la medida en que se aproximan a 1 y -1, la correlación aumenta de forma positiva o negativa respectivamente. Para el conjunto de datos 1 de las Empresas Clientes podemos concluir a partir de la matriz de correlación y dispersión del gráfico 4, que se evidencia correlaciones positivas y negativas poco fuertes, ya que el valor más alto de correlación positiva apenas es de 0.646 y de correlación negativa es de -0.684.

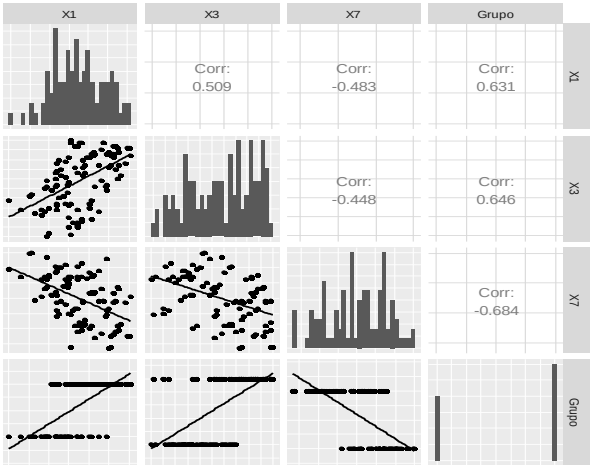

Gráfico 4: Matriz de correlación y dispersión del conjunto de datos 1, Empresas Clientes.

En el gráfico 5 de matriz de dispersión para las variables en estudio se puede distinguir los grupos de clasificación de todo el conjunto de datos, con rojo las empresas que pertenecen al grupo 0 y con verde las que pertenecen al grupo 1. En todos los cruces se puede identificar que hay traslapo pequeño y grupos bien definidos.

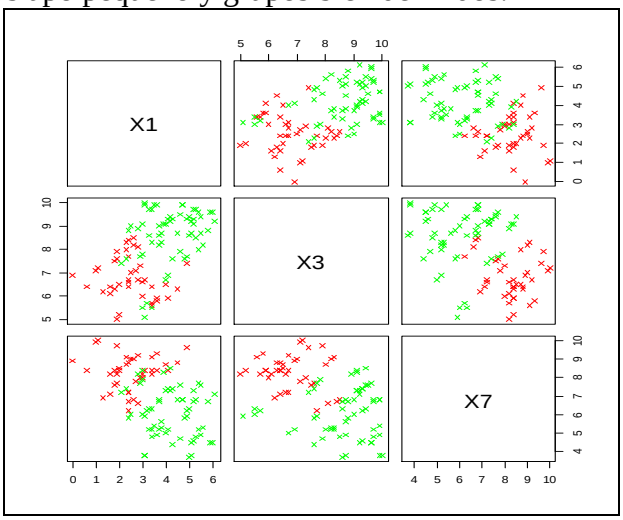

Gráfico 5: Diagramas de dispersión con identificación de los grupos 0 con color rojo y 1 con verde, Empresas Clientes.

Para el conjunto de datos 2, rendimientos financieros bancos japoneses se concluye a partir de la matriz de correlación y dispersión del gráfico 6, que se evidencia correlaciones positivas y negativas muy poco fuertes, ya que el valor más alto de correlación positiva apenas es de 0.594 y de correlación negativa es de -0.584.

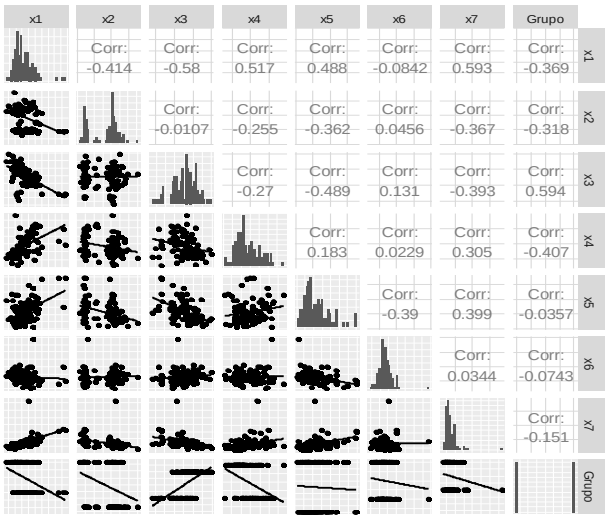

Gráfico 6: Matriz de correlación y dispersión del conjunto de datos 2. Bancos Japoneses.

En el gráfico 7 de matriz de dispersión para las variables en estudio se puede distinguir los grupos de clasificación, con rojo los bancos que pertenecen al grupo 1 y con verde los que pertenecen al grupo 2. Las variables x1, x3, x4 y x7 fueron las que tuvieron correlaciones más altas, se puede observar que hay un traslapo más fuerte y por tanto los grupos se encuentran menos definidos.

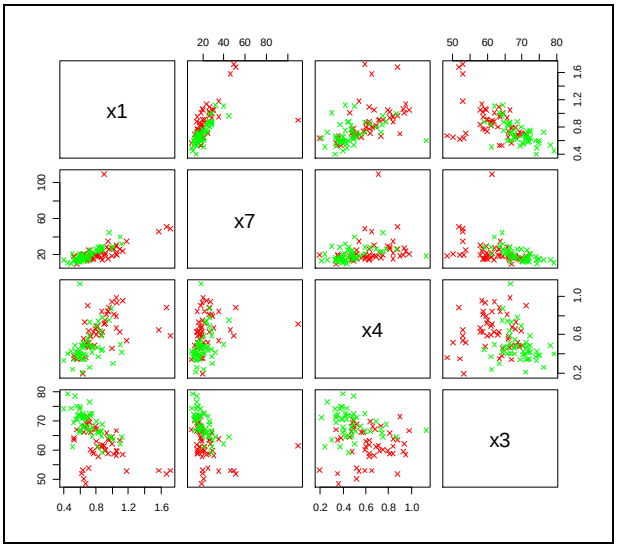

Gráfico 7: Diagramas de dispersión con identificación de los grupos 1 con color rojo y 2 con verde, Rendimientos de bancos Japoneses.

La verificación de los supuestos de Normalidad multivariada de Mardia, de homogeneidad de varianzas y de igualdad de medias, se resumen en la tabla 4, para cada uno de los conjuntos de datos usados en este trabajo.

	Nivel de error o de significancia: 0,05		
Conjunto de datos	**Normalidad Multivariada de Mardia**	**Homogeneidad de varianzas**	**Igualdad de Medias**
Conjunto 1: Empresas clientes	Se cumple	No se cumple	No se cumple
Conjunto 2: Rendimiento bancos japoneses	No se cumple	No se cumple	No se cumple

Tabla 4: Resumen del cumplimiento de los supuestos de normalidad, homogeneidad de varianzas e igualdad de medias, para los dos conjuntos de datos.

En la prueba normal multivariada de Mardia para el conjunto de datos 1 (Empresas clientes), como el valor de P para el sesgo (P= 0.7441263) y el valor de P para la curtosis (P= 0.9983747) ambas son mayores que el nivel de significancia $\alpha = 0.05$, Entonces el conjunto de datos se comporta de forma normal multivariada, es decir $X \sim N_P(\mu, \Sigma)$, o los coeficientes de simetría y curtosis son, respectivamente, $\beta_{1,p} = 0$ y $\beta_{2,p} = p(p+2)$.
Para el conjunto de datos 2 (Bancos japoneses), como el valor de P para el sesgo (P= 1.081015e-261) y el valor de P para la curtosis (P= 0) ambas no son mayores que el nivel de significancia $\alpha = 0.05$, Entonces el conjunto de datos no se comporta de forma normal multivariada. (Ver algoritmo 1, en la página de anexos)

En la prueba M-Box para la verificación del supuesto de homogeneidad de varianzas para el conjunto 1 de datos, como el valor de P (P= 0.0267025) no es mayor que el nivel de significancia $\alpha = 0.05$, se rechaza la hipótesis nula de igualdad de varianzas, en otras palabras no se cumple el supuesto de homogeneidad de varianzas.
Para el conjunto de datos 2, el valor de P (P= 4.837943e-22) no es mayor que el nivel de significancia $\alpha = 0.05$, no se cumple el supuesto de homogeneidad de varianzas. (Ver algoritmo 2, en la página de anexos)

En la prueba de igualdad de medias, para el conjunto 1 de datos (Empresas clientes), como el estadístico de prueba F= 42.46074 es mayor que el valor crítico $F_{(0.05,3,97)} = 2.6983$, se rechaza la hipótesis nula de igualdad de medias. Para el conjunto 2 de datos (Bancos Japoneses), como el estadístico de prueba F= 10.0954 es mayor que el valor crítico $F_{(0.05,7,97)} = 2.109657$, se rechaza la hipótesis nula de igualdad de medias. (Ver algoritmo 3, en la página de anexos)

La tabla 5 muestra, los puntajes y predicciones, sobre los datos de validación (Un total de 40) los cuales se obtuvieron primero de obtener los coeficientes y parámetros estimados con los datos de calibración (un total de 60). Al final se muestra las tasas de error.

Puntajes, Tasas de error, Predicción, Datos de Validación. Conjunto 1 Empresas Clientes										
		AD Lineal		AD Cuadrático		Modelo Almeida			Modelo Sueyoshi	
DMU	Grupo	Fisher		Fisher		F. G. 0	F. G. 1	Criterio Mayor	Etapa 2	
SEL	Original	Disc Score	Predicción	Disc Score	Predicción	hg0	hg1	Predicción	ED*-d	Predicción
3	0	-1,700805057	0	*	0	0,33	<0	0	1,12567	0
4	0	-1,18048387	0	*	0	0,53	<0	0	0,86131	0
10	0	-1,390538457	0	*	0	0,21	<0	0	1,0598	0
27	0	-1,185533316	0	*	0	0,45	<0	0	0,80058	0
30	0	-1,437246807	0	*	0	0,05	<0	0	1,05297	0
34	0	-1,515745919	0	*	0	0,41	<0	0	1,0692	0
35	0	-0,172951857	1	*	1	<0	0,47	1	0,29587	1
37	0	-2,180160113	0	*	0	0,67	<0	0	1,42763	0
40	0	-2,148470529	0	*	0	1,07	<0	0	1,3412	0
41	0	-1,245370796	0	*	0	0,87	<0	0	0,86043	0
57	0	-1,133898258	0	*	0	0,00	<0	0	1,0598	0
60	0	-1,460814117	0	*	0	0,35	<0	0	1,05033	0
75	0	-1,486876855	0	*	0	0,54	<0	0	1,02984	0
83	0	-2,27579154	0	*	0	1,27	<0	0	1,31895	0
85	0	-0,082824539	1	*	1	<0	0,49	1	0,31981	1
87	0	-0,152645535	1	*	1	<0	0,49	1	0,33435	1
94	0	-2,615280689	0	*	0	1,11	<0	0	1,46354	0
98	0	-2,599554952	0	*	0	1,14	<0	0	1,47808	0
9	1	1,054841644	1	*	1	<0	0,44	1	-0,00264	1
16	1	1,851504088	1	*	1	<0	1,66	1	-0,63483	1
18	1	1,024436431	1	*	1	<0	0,75	1	-0,16258	1
19	1	1,537954941	1	*	1	<0	0,87	1	-0,27809	1
21	1	0,523029184	1	*	1	<0	0,77	1	0,04267	1
22	1	1,042044969	1	*	1	<0	1,15	1	-0,26869	1
38	1	0,473629884	1	*	1	<0	0,49	1	0,17184	1
44	1	2,653079044	1	*	1	<0	1,95	1	-1,02741	1
46	1	1,562243418	1	*	1	<0	1,02	1	-0,34139	1
55	1	1,15610728	1	*	1	<0	1,11	1	-0,26869	1
56	1	-0,790188843	0	*	0	0,35	<0	0	0,69792	1
62	1	2,507810205	1	*	1	<0	1,77	1	-0,89141	1
63	1	0,544273226	1	*	1	<0	0,50	1	0,15473	1
64	1	-0,742094943	0	*	1	<0	0,06	1	0,52175	1
66	1	0,58217389	1	*	1	<0	0,68	1	0,07432	1
69	1	2,130773878	1	*	1	<0	1,46	1	-0,70577	1
74	1	0,969294911	1	*	1	<0	0,49	1	-0,00264	1

76	1	0,02361275	1	*	1	<0	0,00	1	0,47123	1
77	1	1,15234904	1	*	1	<0	1,14	1	-0,2952	1
78	1	1,590528248	1	*	1	<0	1,50	1	-0,49369	1
91	1	-0,81659087	0	*	0	0,29	<0	0	0,72957	0
100	1	0,525849705	1	*	1	<0	0,92	1	-0,0044	1
			Tasa Error=		Tasa Error=			Tasa Error=		Tasa Error=
			15,0%		12,5%			12,50%		10%

Tabla 5: Puntajes, Tasas de error, Predicción, Datos de Validación. Conjunto 1 Empresas Clientes

Para el conjunto 1 de datos, que hace referencia a las empresas clientes, se puede ver claramente en la Tabla 5, como el modelo de Discriminación DEA-Extendido o modelo de Sueyoshi, fue el que tuvo una menor tasa de error de clasificación, con solo un 10%, frente al modelo de clasificación por técnica de optimización o modelo de Almeida, el cual tuvo una tasa de error de clasificación del 12.5%, al igual que las funciones cuadráticas de Fisher. El modelo de discriminación que tuvo una mayor tasa de error de clasificación fue el de las funciones lineales de Fisher con 15%, esto sin lugar a dudas debido a la naturaleza multivariada de los datos que no cumplían con el supuesto de homogeneidad de varianzas y por ello era más adecuado estadísticamente usar las funcione cuadráticas de Fisher para la discriminación.

Se aprecia en la Tabla 5, resaltado con azul, que los errores de clasificación en todos los tres modelos a comparar coinciden casi en las mismas DMUS en este caso en el número 85, 87, 56 y 91 principalmente.

El modelo DEA-DA Extendido de Sueyoshi respondió de forma positiva, incluso mejorando las tasas de mala clasificación en el modelo de Almeida, con el conjunto de datos 1, el cual fue inicialmente probado con el modelo de Almeida.

En [4], Almeida usando el conjunto de datos 1, mostró que el modelo de clasificación con solución aproximada por técnica de optimización propuesto por él mismo, mejoraba las tasas de mala clasificación al usar las funciones lineales de Fisher, lo cual se puede evidenciar en la Tabla 5 como cierto. Sin embargo, el autor desconoció la naturaleza multivariada de los datos y al no cumplirse la homogeneidad de varianzas entre los grupos de clasificación, se puede ver en la tabla 5, como el desempeño del modelo de Almeida iguala al desempeño realizado por las funciones Cuadráticas de Fisher, al obtener por igual tasas de error del 12.5%.

En términos generales los tres modelos de discriminación, respondieron de forma positiva, al obtener tasas de mala clasificación bajas, esto debido también al hecho de que el conjunto de datos 1 de Empresas Clientes, tiene 3 variables y se está cumpliendo con el supuesto de normalidad Multivariada.

Coeficientes funciones lineales de Fisher	
x_1	0,4130542
x_3	0,4212776
x_7	-0,5491761

Tabla 6: Coeficientes funciones lineales de Fisher, obtenidos con los datos de calibración, conjunto Empresas Clientes.

Pesos y Parámetros estimados Etapa 1 Modelo Sueyoshi				Pesos y Parámetros estimados Etapa 2 Modelo Sueyoshi			
λ_1^+	0	λ_1	-0,3112	λ_1^+	0,1835	λ_1	-0,1454
λ_2^+	0	λ_2	-0,2542	λ_2^+	0	λ_2	-0,1711
λ_3^+	0,4347	λ_3	0,4347	λ_3^+	0,3165	λ_3	0,3165
λ_1^-	0,3112	$d+1$	1,019	λ_1^-	0,3289	$d+1$	1,1257
λ_2^-	0,2542	d	0,019	λ_2^-	0,1711	d	0,1257
λ_3^-	0			λ_3^-	0	c	0,7074

Tabla 7: Pesos y parámetros estimados en las etapas 1 y 2 Modelo Sueyoshi, conjunto Empresas Clientes

La tabla 6, muestra los coeficientes lineales de Fisher estimados con los datos de calibración del conjunto 1 de Empresas Clientes y a partir de los cuales se obtienen los puntajes para cada una de las muestras del conjunto de validación, los cuales son mostradas en la tabla 5.

La tabla 7, Muestra los pesos y parámetros estimados con el modelo de Sueyoshi, con los datos de calibración, tanto etapa 1 como 2. En la etapa 2 se obtiene el valor de c el cual es un criterio de clasificación usado con los datos de validación y los cuales son mostrados en la Tabla 5.

La tabla 8 muestra, los puntajes y predicciones, sobre los datos de validación (Un total de 40 muestras) los cuales se obtuvieron primero de obtener los coeficientes y parámetros estimados con los datos de calibración (un total de 60). Al final se muestra las tasas de error obtenidas con los tres modelos y usando el conjunto de datos de rendimientos de Bancos Japoneses.

Se puede observar en la tabla 8, que el modelo de discriminación usando las funciones lineales de Fisher fue el que tuvo mejor desempeño, al obtener una tasa de error de solo el 7.5%

	Puntajes, Tasas de error, Predicción, Datos de Validación. Conjunto 2 Bancos Japoneses									
		AD Lineal		AD Cuadrático		Modelo Almeida			Modelo Sueyoshi	
DMU	Grupo	Fisher		Fisher		F. G. 0	F. G. 1	Criterio Mayor	Etapa 2	
SELECC	Original	Disc Score	Predicción	Disc Score	Predicción	hg0	hg1	Predicción	ED*-d	Predicción
3	1	-2,666660785	1	*	1	0	0	*	-6,778227	1
8	1	-2,231676082	1	*	1	0	0	*	-7,943324	1
9	1	-3,058522797	1	*	1	2,28102774	0	1	-7,379318	1
11	1	-1,620051085	1	*	1	0E+00	0	*	-8,475252	1
15	1	-2,255180066	1	*	1	0	0	*	-7,583833	1
18	1	-1,03040961	1	*	1	2,59220077	0	1	-8,312724	1
31	1	-0,469468806	1	*	1	2,9450688	0	1	-8,933568	1
33	1	-1,159485173	1	*	1	0,90839387	0	1	-8,360476	1
35	1	0,52694616	2	*	2	0	0	*	-9,64224	2
38	1	-0,207819083	1	*	1	3,62380929	0	1	-9,320951	1
40	1	-0,458659687	1	*	1	5,24E-16	0	1	-8,878214	1
41	1	0,051612341	1	*	2	1,49960562	0	1	-9,119862	1
47	1	0,533540891	2	*	1	0,84121483	0	1	-9,820581	2
48	1	-0,020399898	1	*	2	0	0,1876263	2	-9,089982	1
49	1	-0,13629895	1	*	2	0,98209014	0	1	-9,109682	1
52	2	0,158280281	1	*	1	0	0	*	-9,217433	1
53	2	0,933480292	2	*	2	0	1,3787641	2	-9,159252	1
54	2	1,169363122	2	*	1	0	0	*	-8,776099	1
56	2	0,629965518	2	*	2	0	0	*	-9,061318	1
60	2	1,047530388	2	*	2	0	0	*	-9,492647	2
64	2	1,32836067	2	*	2	0	0	*	-10,33951	2
66	2	1,274442713	2	*	2	0	0,547461	2	-9,463602	2
68	2	2,025013815	2	*	2	0	2,2711938	2	-9,613572	2
69	2	1,04090266	2	*	2	0	0,0087779	2	-9,724369	2
71	2	1,291266221	2	*	2	0	0,3366716	2	-9,714051	2
72	2	1,406271816	2	*	2	0	2,0455268	2	-9,535941	2
74	2	0,808949058	2	*	2	0	0,3065787	2	-9,484348	2
76	2	1,808053228	2	*	2	0	0	*	-10,07372	2
77	2	1,304854189	2	*	2	0	0,00E+00	*	-10,07364	2
80	2	2,062934873	2	*	2	0	0	*	-10,46329	2
90	2	2,988537716	2	*	2	0	0,00E+00	*	-10,44254	2
91	2	3,157458307	2	*	2	0	0,6798823	2	-10,01507	2
92	2	2,881298011	2	*	2	0	0,7852304	2	-10,13283	2
93	2	3,055964507	2	*	2	0	0	*	-10,73655	2
94	2	3,034135581	2	*	2	0	0,9215934	2	-10,34188	2

95	2	1,267768654	2	*	2	0	0,4599898	2	-9,297424	1
96	2	1,792785819	2	*	1	0	0,9557242	2	-9,698382	2
98	2	2,199500308	2	*	2	0	1,7249752	2	-9,855197	2
99	2	3,34216342	2	*	1	1,2838E+17	0,859255	1	-10,80066	2
100	2	3,803098017	2	*	2	1,2838E+17	0	1	-11,19508	2
			Tasa Error=		Tasa Error=			Tasa Error=		Tasa Error=
			7,5%		20%			*		17.5%

Tabla 8: Puntajes, Tasas Error, Predicción, Conjunto de Validación. Datos Bancos Japoneses.

Las funciones cuadráticas de Fisher, fue el modelo con mayor tasa de error con un 20%. El modelo DEA-DA Extendido de Sueyoshi, tuvo un desempeño aceptable, al obtener una tasa de error del 17.5%

El modelo de Almeida, definitivamente no tuvo un buen desempeño; un poco menos de la mitad de las observaciones no se pudieron clasificar a uno de los dos grupos y por tanto no se pudieron estimar las tasas de error de clasificación. Posiblemente esto debido al gran número de variables del conjunto de datos o a la falta de normalidad multivariada y de homogeneidad de varianzas.

El modelo DEA-DA Extendido de Sueyoshi nuevamente respondió de forma positiva, incluso a pesar de tener una tasa de error aceptable, superando nuevamente el modelo de Almeida que no respondió ante la naturaleza del conjunto 2 de datos de Rendimientos financieros de los bancos Japoneses.

Se aprecia en la Tabla 8, resaltado con azul, que los errores de clasificación con las funciones lineales y cuadráticas de Fisher y con el modelo DEA-DA Extendido de Sueyoshi; coinciden muy poco las DMUS en este caso solo coincidieron las muestras 35, 47, 52 y 54.

Coeficientes funciones lineales de Fisher.	
x_1	-1,174200903
x_2	-0,270945292
x_3	0,195584201
x_4	-2,924735309
x_5	0,36068135
x_6	-0,007194986
x_7	0,003575516

Tabla 9: Coeficientes funciones lineales de Fisher, obtenidos con los datos de calibración, conjunto Rendimientos Bancos Japoneses

Pesos y Parámetros estimados Etapa 1 Modelo DEA-DA Extendido de Sueyoshi				**Pesos y Parámetros estimados Etapa 2 Modelo DEA-DA Extendido de Sueyoshi**			
λ_1^+	0	d	-12,9106	λ_1^+	0	λ_1	0
λ_2^+	0,0321	λ_1	0	λ_2^+	0,0606	λ_2	0,0606
λ_3^+	0	λ_2	0,0321	λ_3^+	0	λ_3	-0,1571
λ_4^+	0,4281	λ_3	-0,1974	λ_4^+	0,5431	λ_4	0,5431
λ_5^+	0	λ_4	0,4281	λ_5^+	0	λ_5	-0,2122
λ_6^+	0,0203	λ_5	-0,3221	λ_6^+	0,0192	λ_6	0,0192
λ_7^+	0	λ_6	0,0203	λ_7^+	0	λ_7	-0,0078
λ_1^-	0	λ_7	-0,0086	λ_1^-	0	c	-9,4463
λ_2^-	0			λ_2^-	0	d	-9,7962
λ_3^-	0,1974			λ_3^-	0,1571		
λ_4^-	0			λ_4^-	0		
λ_5^-	0,3221			λ_5^-	0,2122		
λ_6^-	0			λ_6^-	0		
λ_7^-	0,0086			λ_7^-	0,0078		

Tabla 10: Pesos y parámetros estimados en las etapas 1 y 2 Modelo Sueyoshi, conjunto Rendimientos Bancos Japoneses

La tabla 9, muestra los coeficientes lineales de Fisher estimados con los datos de calibración del conjunto 2 de Rendimientos de Bancos Japoneses y a partir de los cuales se obtienen los puntajes para cada una de las muestras del conjunto de validación, los cuales son mostradas en la tabla 8.

La tabla 10, Muestra los pesos y parámetros estimados con el modelo de Sueyoshi, con los datos de calibración, tanto etapa 1 como 2. En la etapa 2 se obtiene el valor de c el cual es un criterio de clasificación usado con los datos de validación y los cuales son mostrados en la Tabla 8.

Finalmente los gráficos 8, 9, 10 y 11, corresponden a una visualización de la separación de planos con las funciones lineales y cuadráticas de Fisher para los dos conjuntos de datos.
Para el conjunto 1 de Empresas Clientes, se graficaron las variables de a pares, todos contra todos, dado que el conjunto de datos solo tiene 3 variables.
Para el conjunto 2 de Bancos Japoneses, se graficaron de a pares solo las variables x1, x3, x4 y x7.

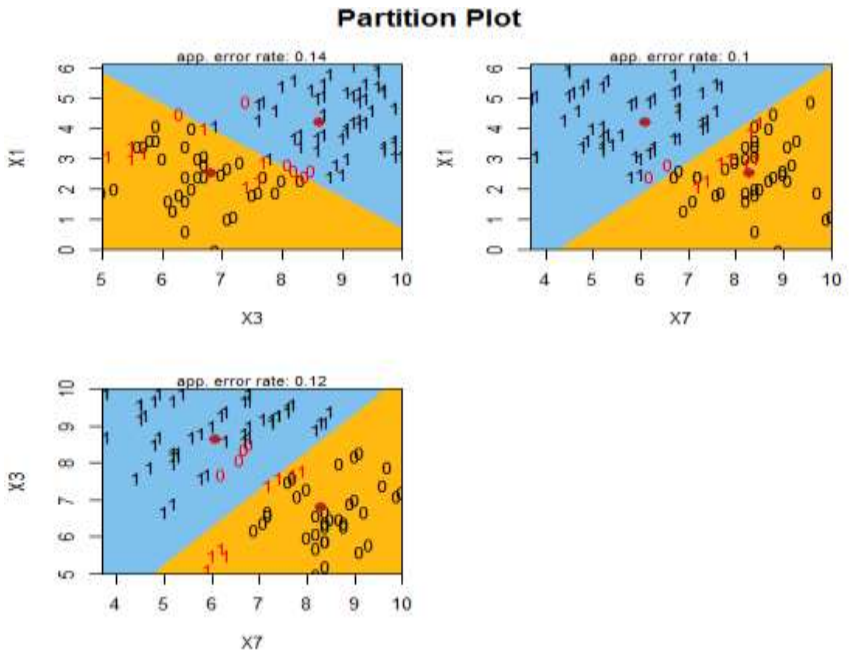

Gráfico 8: Separación de planos con Funciones lineales de Fisher, conjunto de datos 1 Empresas Clientes, x1 vs x3, x1 vs x7 y x3 vs x7

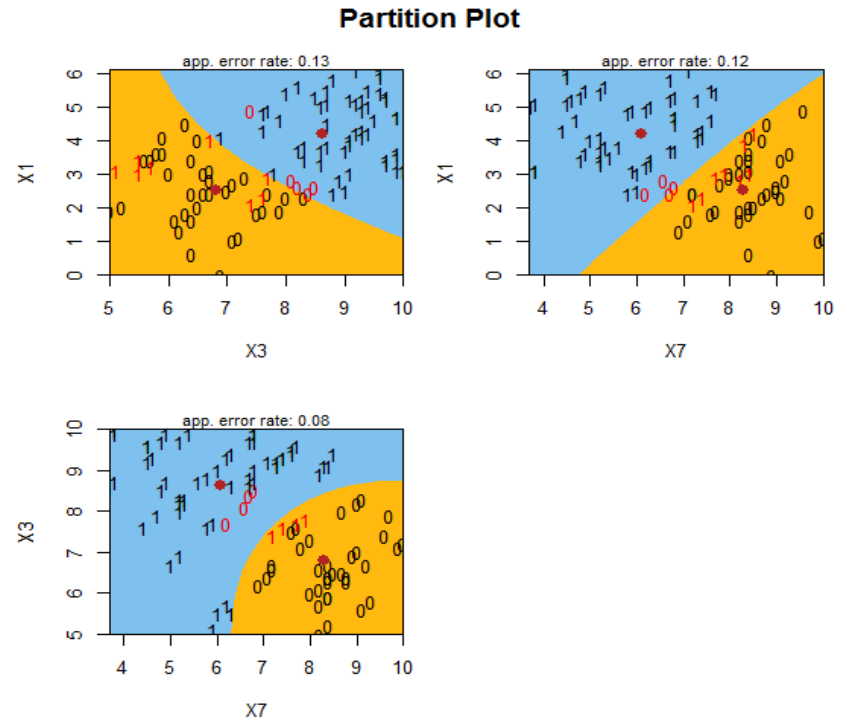

Gráfico 9: Separación de planos con Funciones cuadráticas de Fisher, conjunto de datos 1 Empresas Clientes, x1 vs x3, x1 vs x7 y x3 vs x7

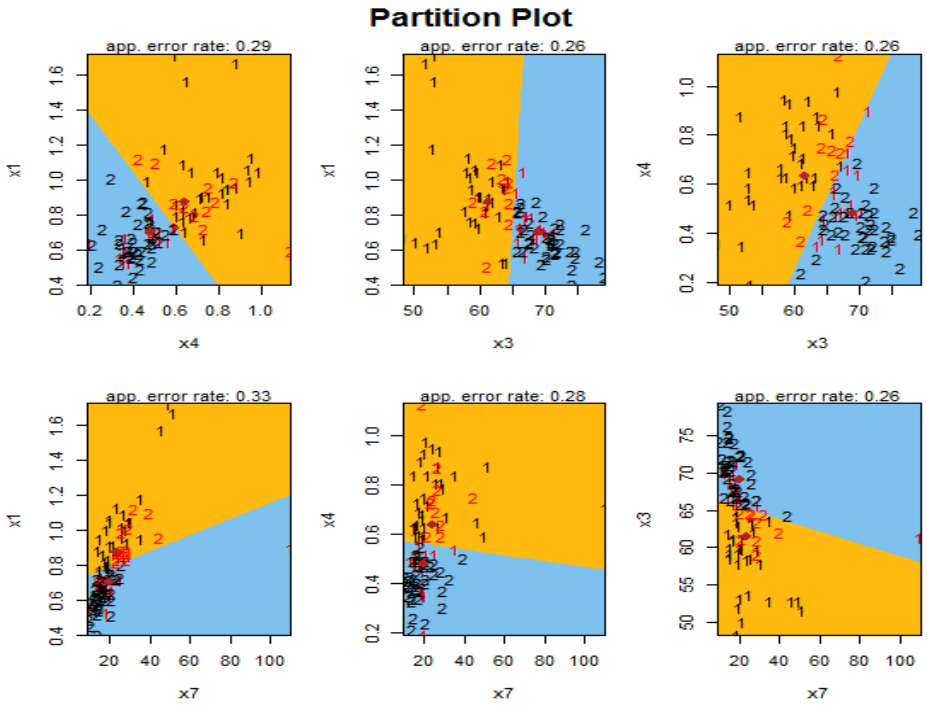

Gráfico 10: Separación de planos con Funciones lineales de Fisher, conjunto de datos 2 Bancos Japoneses, x1 vs x3, x1 vs x7, x1 vs x4, x4 vs x3, x7 vs x4 y x3 vs x7

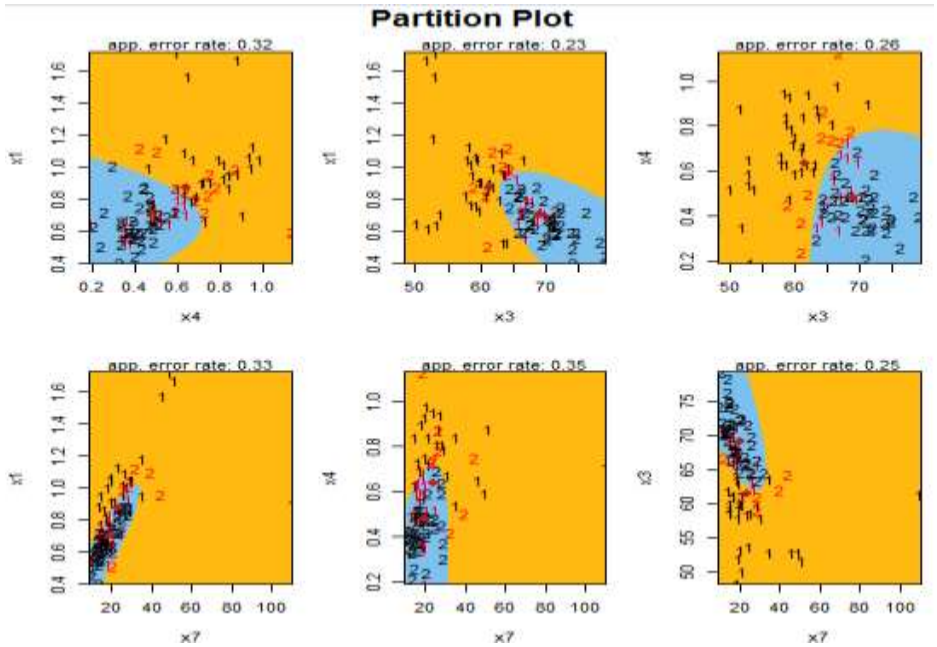

Gráfico 11: Separación de planos con Funciones cuadráticas de Fisher, conjunto de datos 2 Bancos Japoneses, x1 vs x3, x1 vs x7, x1 vs x4, x4 vs x3, x7 vs x4 y x3 vs x7

8. CONCLUSIONES

Se comparó el desempeño de tres modelos para el análisis discriminante: 1) El enfoque DEA - DA extendido, propuestos por Sueyoshi en [16] y [17], 2) El modelo de clasificación con solución aproximada por técnica de optimización propuesto por Almeida en [4] y 3) El modelo clásico de discriminación de Fisher, siguiendo a [3], [5], [6], [10], [11] y [27]. La comparación se realizó al nivel de clasificación para dos grupos. En la literatura no se ha reportado la comparación de estos tres modelos.

Se utilizó dos conjuntos de datos reales propuestos por los mismos autores en sus diferentes artículos. El conjunto 1 hace referencia a información sobre 100 Empresas Clientes y el conjunto 2 sobre los rendimientos financieros de 100 bancos japoneses.

La comparación está basada en el cálculo de una sola tasa de error de clasificación, la cual fue obtenida por el método de Partición de la muestra. Cada conjunto de datos fue dividido en dos partes: Del 100% de los datos se tomó de forma aleatoria el 60% para conformar el conjunto de calibración o entrenamiento. Con el 40% de los datos restantes se conformó el conjunto de validación.

El desempeño del modelo de discriminación con las funciones lineales de Fisher, al usar el conjunto de datos 1 de Empresas Clientes, no fue el mejor, debido a que la tasa de error de clasificación (15%) estuvo por encima de las tasas de los otros modelos. De todas maneras una tasa de error de clasificación del 15% es apenas aceptable. Con el conjunto de datos 2 de Bancos Japoneses, el desempeño de las funciones lineales de Fisher fue el mejor con respecto al de los otros modelos, ya que la tasa de error de clasificación fue del 7.5%, el cual estuvo por debajo de las otras tasas.

El desempeño del modelo de discriminación con las funciones cuadráticas de Fisher, al usar el conjunto de datos 1 de Empresas Clientes, fue el segundo mejor al obtener una tasa de error de clasificación del 12.5% e igualó el desempeño del modelo de Almeida. Con el conjunto de datos 2 de Bancos Japoneses, el desempeño no fue el mejor ya que su tasa de error de clasificación del 20% estuvo por encima de los demás modelos, con excepción del modelo de Almeida cuya tasa no pudo ser estimada por incapacidad del modelo de clasificar las observaciones a uno de los dos grupos. Sin embargo una tasa de error del 20% es apenas aceptable.

De forma general el modelo de Almeida con el conjunto de datos 1 de Empresas clientes igualó el desempeño de las funciones cuadráticas de Fisher con una tasa del 12.5%, pero con el conjunto de datos 2 de Bancos Japoneses el modelo no pudo clasificar un poco menos de la mitad de las observaciones del conjunto de calibración y de validación y por tal motivo no se estimó la tasa de error de clasificación, convirtiendo el modelo de Almeida en el de más bajo desempeño de los tres modelos.

El modelo DEA-DA Extendido de Sueyoshi con el conjunto de datos 1 de Empresas clientes fue el que tuvo mejor tasa de error de clasificación con un 10%, valor que estuvo por debajo de las tasas de los otros modelos. Con el conjunto de datos 2 de Bancos Japoneses, la tasa de error de clasificación con el modelo de Sueyoshi fue del 17.5%, un valor aceptable y estuvo por debajo de las tasas de los modelos de Almeida (cuya tasa no fue estimada) y de las funciones cuadráticas de Fisher (20%), siendo el modelo de Sueyoshi superado solo por el modelo con las funciones lineales de Fisher, el cual tuvo una tasa del 7.5%. Se convierte el modelo DEA-DA Extendido como el de mejor desempeño de los tres modelos.

El modelo DEA-DA Extendido tiene una ventaja con respecto a los otros dos modelos usados en este trabajo y es que es el único modelo que realiza una identificación de los traslapos. Los traslapos en un problema de discriminación son la mayor fuente de mala clasificación y por ello es importante prestar mucha atención y el modelo de Sueyoshi lo hace.

9. RECOMENDACIONES

Comparar metodologías para el análisis de los problemas sobre discriminación deja muchos problemas abiertos, los cuales en su mayoría ya se encuentran siendo abordados actualmente. Sin embargo se expresan algunas de las situaciones que pueden ser consideradas en futuros trabajos con el objetivo de mejorar la metodología de comparación entre los distintos modelos de análisis discriminantes y muy posiblemente reducir las tasas de error de clasificación.
Entre dichas situaciones tenemos por ejemplo el de selección de variables. Este trabajo compara tres modelos de discriminación y para ello usó dos conjuntos de datos. El conjunto 1 de Empresas Clientes, tenía un número muy reducido de variables, mientras que el conjunto 2 de Bancos Japoneses, tenía 7 variables y se trabajó con todas ellas. Esto nos lleva a un nuevo interrogante ¿Existe un número reducido de variables para el conjunto 2, que me permita tener una menor tasa de error de clasificación?, ¿Cuáles variables tienen mayor poder de discriminación?, ¿Cuál método me permite hacer una adecuada selección de variables?

En el momento de implementar un modelo a datos del mundo real, es muy importante evaluar los tiempos computacionales. Los modelos discriminantes con las funciones lineales y cuadráticas de Fisher, se encuentran implementadas en casi todos los Software que existen y son computacionalmente muy rápidos, fácil de implementar e interpretar, razón por las cual son de un uso muy popular y en este aspecto llevan una ventaja muy grande sobre los modelos de análisis discriminantes con enfoque DEA. Particularmente en este trabajo se encontró que el modelo de Almeida es muy extenso y eso lo hace impráctico en el momento de implementarlo, el reto es la construcción de algoritmos o interfaces que permitan la implementación de estos modelos de una forma más accesible.

Este trabajo usó una sola tasa de error de clasificación para comparar el desempeño de los tres modelos, por ello es importante usar más criterios o tasas de error de clasificación ya sea de tipo práctico o empírico.

Otros aspectos importantes a considerar cuando se comparan modelos de discriminación y pueden ser tenidos en cuenta son: el tipo de datos a usar, ya sea económicos, académicos, mercadotecnia, identificación de Outliers, independencia de los parámetros del modelo frente al tamaño de población de cada grupo o mejoramiento de las fronteras de eficiencia.

10. BIBLIOGRAFÍA

[1] Bal, H., Orkcu, H., 2011. A new mathematical programming approach to multi-group classification problems. Computer & Operations Research 38, 105-111.

[2] Boudaghi, E., Farzipoor, R., 2015. Developing a model for determining optimal η in DEA-Discriminant analysis for predicting suppliers' group membership in supply chain. OPSEARCH 52, 134-155.

[3] Clavijo M., Jairo A., Apuntes de Análisis Multivariado, Universidad del Tolima, Ibagué**.**

[4] De Almeida, Henry R., 2006. Um Modelo de Classificação com Solução Aproximada por Técnica de Optimização, Invetigação Operacional, Vol. 26, 179-200.

[5] Díaz M., Luis G., 2002. Estadística Multivariada: Inferencia y Métodos, Universidad Nacional de Colombia, Bogotá.

[6] Hair, J.F.J., Anderson R.E., Tatham, L.T., Black, W.C., Multivariate Data Analysis, Prentice Hall, Upper Saddle River. N.J.

[7] Hernández, F., Correa, J., 2009. Comparación entre tres técnicas de clasificación. Revista Colombiana de Estadística. Vol. 32. 247-265.

[8] Pastor, T. J., Ruiz, L. J. and Sirvent, I., 1999. A Statistical Test for Detecting Influetial Observations in DEA, European Journal of Operational Research 115. 542-554.

[9] Peña, D., 1998. Estadística Modelos y Métodos. Fundamentos, Alianza Universitaria Textos, Madrid.

[10] Pérez, César, Técnicas de Análisis Multivariante de Datos con SPSS.

[11] Rencher, A., 1998. Multivariate Statistical Inference and Applications, New York.

[12] Rezai, F., Saghaei, M y Hosseinzadeh, F., 2010. DEA-Discriminant analysis for three groups. Applied Mathematical Sciences. Vol 4, 3575-3588.

[13] Shang, D., Chun, Y., 2008. An approach for the two group discriminant analysis: An application of DEA. Mathematical and Computer Modelling 47. 970-981.

[14] Soto M., Jose A., Arenas, V., Wilson, 2010. Análisis Envolvente de Datos: De la Teoría a la Práctica, Universidad Tecnológica de Pereira, Pereira.

[15] Sueyoshi, T., 1990. Special algorithm for an additive model in data envelopment analysis. Journal of the Operational Research Society 41, 249-257.

[16] Sueyoshi, T., 1999. DEA-discriminant analysis in the view of goal programming. European Journal of operational Research 115, 564-582.

[17] Sueyoshi, T., 2001. Extended DEA-Discriminant Analysis, European Journal of Operational Research, Vol. 131, 324–351.

[18] Sueyoshi, T., Hwang, S.N., 2004. A use of nonparametric test for DEA-DA: A methodological comparison. Asia-Pacific Journal of Operational Research, Vol. 21, 179-197.

[19] Sueyoshi, T., 2004. Mixed Integer Programming Approach of extended-discriminant analysis, European Journal of Operational Research, Vol. 152, 45–55.

[20] Sueyoshi, T., 2005. Financial ratio analysis of electric power industry. Asia-Pacific Journal of Operational Research, VOL 22, 349-376.

[21] Sueyoshi, T., 2005. A methodological Comparison between standard and two stage mixed integer approaches for discriminant analysis. Asia-Pacific Journal of Operational Research, Vol 22, 513-528.

[22] Sueyoshi, T., 2006. DEA-Discriminant Analysis: Methodological comparison among eight discriminant analysis approaches, European Journal of Operational Research 169, 247–272.

[23] Sueyoshi, T., Goto, M., 2009. Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA-discriminant analysis. European Journal of Operational Research 196, 289-311.

[24] Sueyoshi, T., Goto, M., 2009. Methodological comparison between DEA (data envelopment analysis) and DEA-DA (discriminant analysis) from the perspective of bankruptcy assessment. Eurpean Journal of Operational Research 199, 561-575.

[25] Sueyoshi, T., Goto, M., 2011. A combined use of DEA (Data envelopment analysis) with strong complementary slackness condition and DEA-DA (discriminant analysis). Applied Mathematics Letters 24, 1051-1056.

[26] Tavassoli, M., Reza, G., Farzipoor, R., 2015. Ranking electricity distributions units using slacks-based measure, strong complementary slackness condition and discriminant analysis. Electrical Power and Energy Systems 64, 1214-1220.

[27] Uriel, Ezequiel y Aldás, Joaquín, Análisis Multivariante Aplicado.

11. ANEXOS

11.1 Algoritmo 1. Prueba de Multinormalidad de Mardia

```
# RUTINA R PARA LA PRUEBA DE MULTINORMALIDAD DE MARDIA
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#Y<-read.table("compra.txt",dec=",",header=T)
Y<-read.table("japoneses.txt",dec=",",header=T)
attach(Y)
Y<-subset(Y,select=-Grupo)
Y<-t(Y)
Y<-t(Y)
dimension<-dim(Y)
N<-dimension[1]
P<-dimension[2]
GL.CHI<-(P)*(P+1)*(P+2)/6
j<-matrix(1,N,N)
I<-diag(1,N)
Q<-I-(1/N)*j
S<-(1/(N-1))*(t(Y))%*%Q%*%Y
INV.S<-solve(S)
G.MATRIZ<-Q%*%Y%*%INV.S%*%t(Y)%*%Q
b_1<-(sum(G.MATRIZ*G.MATRIZ*G.MATRIZ))/(N*N)
b_2<-sum(diag(G.MATRIZ*G.MATRIZ))/N
EST_b_1<-N*b_1/6
EST_b_1
EST_b_2<-(b_2-P*(P+2))/sqrt(8*P*(P+2)/N)
EST_b_2
PVAL_ses<-pchisq(EST_b_1,GL.CHI,lower.tail=FALSE)
PVAL_ses
PVAL_cur<-pnorm(EST_b_2,lower.tail=FALSE)
PVAL_cur
if (PVAL_ses>0.05 & PVAL_cur>0.05) cat("Se cumple el supuesto de
Normalidad multivariada \n") else ("No se cumple el supuesto de
normalidad")
```

11.2 Algoritmo 2. Prueba de Homogeneidad de varianzas para dos grupos de M-cox.

```
# PRUEBA DE HOMOGENEIDAD DE DOS VARIANZAS M-cox PARA DOS GRUPOS
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#Y<-read.table("compra.txt",dec=",",header=T)
```

```
Y<-read.table("japoneses.txt",dec=",",header=T)
dimension<-dim(Y)
N<-dimension[1]
colclas<-dimension[2]
n1<-0
for (i in 1:N){
if (Y[i,colclas]==Y[1,colclas]){
n1<-n1+1
}
n1
}
n2<-N-n1
Y<-Y[,-length(Y)]
G1<-Y[1:n1,]
G2<-Y[n1+1:n2,]
q<-2
p<-length(Y)
v<-N-q
v1<-n1-1
v2<-n2-1
S1<-cov(G1)
S2<-cov(G2)
SP<-(1/(N-q))*(((n1-1)*S1)+((n2-1)*S2))
rho<-1-((2+p^2+3*p-1)/(6*(p+1)*(q-1)))*((1/v1)+(1/v2)-(1/v))
rho
z<-v*log(det(SP))-(v1*log(det(S1))+v2*log(det(S2)))
z
EST_phi<-z*rho
EST_phi
gl<-p*(p+1)*(q-1)/2
gl
VALP_phi<-pchisq(EST_phi,gl,lower.tail=FALSE)
VALP_phi
if (VALP_phi>0.05) cat("Se acepta Ho las varianzas son iguales \n")
else cat("Se rechaza Ho las varianzas no son iguales \n")
```

11.3 Algoritmo 3. Comparación de la media de dos poblaciones asumiendo varianzas diferentes.

```
rm(list=ls(all=TRUE))
```

```
setwd("D:/Desktop/datostesis")
Y<-read.table("compra.txt",dec=",",header=TRUE)
#Y<-read.table("japoneses.txt",dec=",",header=TRUE)
dimension<-dim(Y)
n<-dimension[1]
p<-dimension[2]-1
colclas<-dimension[2]
n1<-0
for (i in 1:n){
if (Y[i,colclas]==Y[1,colclas]){
n1<-n1+1
}
n1
}
n2<-n-n1
Y<-Y[,-length(Y)]
G1<-Y[1:n1,]
G2<-Y[n1+1:n2,]
d<-matrix(0,n1,2)
for (i in 1:n1){
d[i,1]<-G1[i,1]-G2[i,1]
d[i,2]<-G1[i,2]-G2[i,2]
d
}
dmedia<-apply(d,2,mean)
dmedia<-t(dmedia)
sdif<-cov(d)
T2<-n1%*%dmedia%*%solve(sdif)%*%t(dmedia)
ESTADISTICO<-(T2/(n1-1))*((n1-p)/p)
VAL_CRITICO<-qf(0.05,p,n-p,lower.tail=FALSE)
VAL_CRITICO
if (ESTADISTICO>VAL_CRITICO) cat("Se rechaza Ho las medias no son
iguales \n") else cat("Se acepta Ho las medias son iguales \n")
```

11.4 Algoritmo 4. Comparación de la media de dos poblaciones asumiendo varianzas iguales.

```
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
Y<-read.table("compra.txt",dec=",",header=TRUE)
```

```
#Y<-read.table("japoneses.txt",dec=",",header=TRUE)
dimension<-dim(Y)
n<-dimension[1]
colclas<-dimension[2]
n1<-0
for (i in 1:n){
if (Y[i,colclas]==Y[1,colclas]){
n1<-n1+1
}
n1
}
n2<-n-n1
Y<-Y[,-length(Y)]
G1<-Y[1:n1,]
G2<-Y[n1+1:n2,]
p<-length(Y)
v<-n1+n2-2
S1<-cov(G1)
S2<-cov(G2)
SP<-(1/(v))*(((n1-1)*S1)+((n2-1)*S2))
mediaG1<-apply(G1,2,mean)
mediaG1<-t(mediaG1)
mediaG1<-t(mediaG1)
mediaG2<-apply(G2,2,mean)
mediaG2<-t(mediaG2)
mediaG2<-t(mediaG2)
EST_T2<-((n1*n2)/(n1+n2))*t((mediaG1-
mediaG2))%*%solve(SP)%*%(mediaG1-mediaG2)
F<-qf(0.05,p,v-p+1,lower.tail=FALSE)
VAL_CRIT_F<-((v*p)/(v-p+1))*F
if (EST_T2>VAL_CRIT_F) cat("Se rechaza Ho las medias no son iguales
\n") else cat("Se acepta Ho las medias son iguales \n")
```

11.5 Algoritmo 5. Análisis exploratorio de los dos conjuntos de datos

```
setwd("D:/Desktop/datostesis")
rm(list=ls(all=TRUE))
#datos<-read.table("compra.txt",dec=",", header=TRUE)
datos<-read.table("japoneses.txt",dec=",", header=TRUE)
```

```
# Analisis exploratorio de los datos
#-------------------------------------------------------------
library(dplyr)
require(GGally)
ggpairs(datos,
        lower = list(continuous = "smooth"),
        diag = list(continuous = "bar"),
        axisLabels = "none")

par(mfrow= c(1,1))
# Asociación entre variables cuantitativas más correlacionadas.
# Se colorean los casos en función del rendimiento
pairs(x = datos[, c("x1", "x7", "x4","x3")],              #tener
cuidado
      col= ifelse(datos$Grupo==1, "red", "green"),        #tener
cuidado
      pch = 4)
```

11.6 Algoritmo 6. Análisis Discriminante con las Funciones lineales de Fisher.

```
library(MASS)
datos.lda = lda(Grupo~.,datos)
attributes(datos.lda)
datos.lda$prior     #probabilidades a priori
datos.lda$counts    #tamaño muestral por grupo
datos.lda$means     #promedios por grupo y variables
datos.lda$scaling   #coeficientes de los discriminantes lineales
datos.lda$lev       #Nombre de cada nivel de cada grupo
datos.lda$svd       #el poder discriminate de cada función
datos.lda$N         #número total de datos
datos.lda$call      #llamado de la fórmula
datos.lda$terms     #términos de la fórmula
datos.lda$xlevels
plot(datos.lda)
```

11.7 Algoritmo 7. Método de Resustitución, Tasa de Error Aparente para las funciones lineales de Fisher.

```
nuevo<-datos[,-length(datos)]
predic.lda.resust<-predict(datos.lda,newdata=nuevo)$class
tabla.resustitucion<-table(datos[,length(datos)],predic.lda.resust)
tabla.resustitucion
tasa.resustitucion<- 1-sum(predic.lda.resust ==
datos[,length(datos)])/datos.lda$N
tasa.resustitucion
```

11.8 Algoritmo 8. Método de Validación Cruzada para las funciones lineales de Fisher.

```
library(MASS)
predic.lda.CV = lda(Grupo~.,datos, CV=T)$class
nuevo<-datos[,-length(datos)]
tabla.CV<-table(datos[,length(datos)],predic.lda.CV)
tabla.CV
tasa.CV<- 1-sum(predic.lda.CV == datos[,length(datos)])/datos.lda$N
tasa.CV
```

11.9 Algoritmo 9. Método de Partición de la muestra, muestra de entrenamiento y de validación para las funciones lineales de Fisher.

```
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#datos<-read.table("compra.txt",dec=",", header=TRUE)
datos<-read.table("japoneses.txt",dec=",", header=TRUE)

library(MASS)
set.seed(1)
n1<-floor(nrow(datos)*0.7)                # n1=tamaño muestra
entrenamiento
(datos.indices <- sample(1:nrow(datos),size=n1))
(datos.entrenamiento <- datos[datos.indices,])
(nuevo <- datos[-datos.indices,])         #muestra validacion
dim(nuevo)
(datos <- datos.entrenamiento)            #muestra entrenamiento
dim(datos)
```

```
(datos.lda = lda(Grupo~.,datos))
(datos.lda$N)
(prediccion.lda<-predict(datos.lda,newdata=nuevo))
(prediccion.particion<-predict(datos.lda,newdata=nuevo)$class)
(tabla.particion<-
table(nuevo[,length(datos)],prediccion.particion))
(tasa.particion<- 1-sum(predict(datos.lda,newdata=nuevo)$class ==
nuevo[,length(datos)])/sum(tabla.particion))
```

11.10 Algoritmo 10. Análisis Discriminante Cuadrático y cálculo de la tasa de error aparente.

```
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#datos<-read.table("compra.txt",dec=",", header=TRUE)
datos<-read.table("japoneses.txt",dec=",", header=TRUE)

library(MASS)
nuevo<-datos[,-length(datos)]
(datos.qda = qda(Grupo~.,datos))
(attributes(datos.qda))
datos.qda$prior     #probabilidades a priori
datos.qda$counts    #tamaño muestral por grupo
datos.qda$means     #promedios por grupo y variables
datos.qda$scaling   #coeficientes de los discriminantes lineales
datos.qda$lev       #Nombre de cada nivel de cada grupo
datos.qda$svd       #el poder discriminate de cada función
datos.qda$N         #número total de datos
datos.qda$call      #llamado de la fórmula
datos.qda$terms     #términos de la fórmula
datos.qda$xlevels   #
(prediccion.qda<-predict(datos.qda,newdata=nuevo))
(tabla.qda<-table(datos[,length(datos)],prediccion.qda$class))
(tasa.qda<- 1-sum(predict(datos.qda,nuevo)$class ==
datos[,length(datos)])/datos.qda$N)
```

11.11 Algoritmo 11. Método de validación cruzada "Funciones cuadráticas de Fisher"

```
library(MASS)
predic.qda.CV = qda(Grupo~.,datos, CV=T)$class
nuevo<-datos[,-length(datos)]
tabla.CV<-table(datos[,length(datos)],predic.qda.CV)
tabla.CV
tasa.CV<- 1-sum(predic.qda.CV == datos[,length(datos)])/datos.qda$N
tasa.CV
```

11.12 Algoritmo 12. Método de Partición de la muestra con Funciones cuadráticas de Fisher.

```
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#datos<-read.table("compra.txt",dec=",", header=TRUE)
datos<-read.table("japoneses.txt",dec=",", header=TRUE)

library(MASS)
set.seed(1)
n1<-floor(nrow(datos)*0.7)                # n1=tamaño muestra
entrenamiento
(datos.indices <- sample(1:nrow(datos),size=n1))
(datos.entrenamiento <- datos[datos.indices,])
(nuevo <- datos[-datos.indices,])         #muestra validacion
dim(nuevo)
(datos <- datos.entrenamiento)            #muestra entrenamiento
dim(datos)

(datos.qda = qda(Grupo~.,datos))
(datos.qda$N)
(prediccion.qda<-predict(datos.qda,newdata=nuevo))
(prediccion.particion<-predict(datos.qda,newdata=nuevo)$class)
(tabla.particion<-
table(nuevo[,length(datos)],prediccion.particion))
(tasa.particion<- 1-sum(predict(datos.qda,newdata=nuevo)$class ==
nuevo[,length(datos)])/sum(tabla.particion))
```

11.13 Algoritmo 13. Visualización de los modelos lineales y cuadráticos de Fisher.

```
rm(list=ls(all=TRUE))
setwd("D:/Desktop/datostesis")
#datos<-read.table("compra.txt",dec=",", header=TRUE)
datos<-read.table("japoneses.txt",dec=",", header=TRUE)
library(MASS)
library(klaR)
# Representación del LDA respecto a los datos de test
partimat(formula = as.factor(Grupo) ~ x1+x4+x3+x7, data = datos,
method = "lda", prec = 400, image.colors = c("darkgoldenrod1",
"skyblue2"), col.mean = "firebrick", nplots.vert = 2)
partimat(formula = as.factor(Grupo) ~ x1+x4+x3+x7, data = datos,
method = "qda", prec = 400, image.colors = c("darkgoldenrod1",
"skyblue2"), col.mean = "firebrick", nplots.vert = 2)
```

11.14 Algoritmo 14. Rutina para la obtención de los valores de hg, hgo y tasas de error de clasificación por el método de Partición de la muestra (datos de calibración y validación) en el modelo de Almeida, con el conjunto 1 de datos de Empresas Clientes.

```
setwd("D:/Desktop/datostesis")
rm(list=ls(all=TRUE))
g0.orig.cal<-read.table("compracal0.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
max_x1<-max(x1)
max_x3<-max(x3)
min_x7<-min(x7)
x1a<-max_x1-x1
x3a<-max_x3-x3
x7a<-x7-min_x7
g0.cal<-rbind(x1a,x3a,x7a)
g0.cal<-t(g0.cal)

dimension<-dim(g0.cal)
n1<-dimension[1]

library(boot)
h<-numeric(n1)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-t(g0.cal)
vectorTermIndep <- matrizRestriccionesMenorOIgual[,1]
```

```
lambda<-matrix(1,1,n1)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(a=h, A1=matrizRestriccionesMenorOIgual,
b1=vectorTermIndep, A3=lambda, b3=lambdaTermInd,maxi=TRUE)
hmo<-mySimplex$value
hmo

hmgo<-((1-hmo)*hmo)/((1-hmo)*hmo)
hmgo

ho<-numeric(n1-1)
for (i in 2:(n1)){
matrizRestriccionesMenorOIgual[,1]<-
matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,n1)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=matrizRestriccionesMenorOIgual,
b1=vectorTermIndep, A3=lambda, b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
}
ho
ho[1]<-hmo
hgo<-numeric(n1)
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo

# ------------------------------------------------------------------
# obtención de las fronteras del grupo 0 del conjunto de
calibración
# ------------------------------------------------------------------
p<-numeric(n1)
for (i in 1:n1){
if (ho[i]==0 & hgo[i]==0){
p[i]<-i
}else{
}
p
}
p

g0.orig.cal<-read.table("compracal0.txt",dec=",", header=TRUE)
```

```
front0<-
rbind(g0.orig.cal[1,],g0.orig.cal[5,],g0.orig.cal[12,],g0.orig.cal[
13,],g0.orig.cal[17,])     #tener cuidado el 1 no cambia
attach(front0)
max_x1<-max(x1)
max_x3<-max(x3)
min_x7<-min(x7)
x1a<-max_x1-x1
x3a<-max_x3-x3
x7a<-x7-min_x7
f0<-rbind(x1a,x3a,x7a)
f0

#-----------------------------------------------------------------
# valores de h y hg con las fronteras originaes grupo 0, para los
dos grupos calibracion
# -----------------------------------------------------------------
setwd("D:/Desktop/datostesis")
g0.orig.cal<-read.table("compracal.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
x1a<-max_x1-X1
x3a<-max_x3-X3
x7a<-X7-min_x7
g0.cal<-rbind(x1a,x3a,x7a)

dimension<-dim(g0.cal)
n1<-dimension[2]

dimension<-dim(f0)
nf<-dimension[2]

library(boot)
h<-numeric(nf)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-g0.cal

ho<-numeric(n1)
for (i in 1:(n1)){
f0[,1]<-matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,nf)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=f0, b1=vectorTermIndep, A3=lambda,
b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
```

```
}
ho

hgo<-numeric(n1)
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo
hgocal<-hgo
write.table(hgo, 'hgocal.txt', sep='\t', dec=',')

#-----------------------------------------------------------------
ahora lo mismo con validacion
#-----------------------------------------------------------------
setwd("D:/Desktop/datostesis")
g0.orig.cal<-read.table("compraval.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
x1a<-max_x1-X1                  #tener cuidado
x3a<-max_x3-X3                  #tener cuidado
x7a<-X7-min_x7                  #tener cuidado
g0.cal<-rbind(x1a,x3a,x7a)

dimension<-dim(g0.cal)
n1<-dimension[2]
dimension<-dim(f0)
nf<-dimension[2]

library(boot)
h<-numeric(nf)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-g0.cal

ho<-numeric(n1)
for (i in 1:(n1)){
f0[,1]<-matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,nf)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=f0, b1=vectorTermIndep, A3=lambda,
b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
}
ho

hgo<-numeric(n1)
```

```
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo
hgoval<-hgo
write.table(hgo, 'hgoval.txt', sep='\t', dec=',')

#-----------------------------------------------------------------
#-----------------------------------------------------------------

# Ajuste de variables de los datos de calibración del grupo 1
# y obtención de la medida ho y hgo
# -----------------------------------------------------------------
setwd("D:/Desktop/datostesis")
#rm(list=ls(all=TRUE))
g0.orig.cal<-read.table("compracal1.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
min_x1<-min(x1)
min_x3<-min(x3)
max_x7<-max(x7)
x1a<-x1-min_x1
x3a<-x3-min_x3
x7a<-max_x7-x7
g0.cal<-rbind(x1a,x3a,x7a)
g0.cal<-t(g0.cal)

dimension<-dim(g0.cal)
n1<-dimension[1]

library(boot)

h<-numeric(n1)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-t(g0.cal)
vectorTermIndep <- matrizRestriccionesMenorOIgual[,1]
lambda<-matrix(1,1,n1)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(a=h, A1=matrizRestriccionesMenorOIgual,
b1=vectorTermIndep, A3=lambda, b3=lambdaTermInd,maxi=TRUE)
hmo<-mySimplex$value
hmo

hmgo<-((1-hmo)*hmo)/((1-hmo)*hmo)
hmgo
```

```
ho<-numeric(n1-1)
for (i in 2:(n1)){
matrizRestriccionesMenorOIgual[,1]<-
matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,n1)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=matrizRestriccionesMenorOIgual,
b1=vectorTermIndep, A3=lambda, b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
}
ho
ho[1]<-hmo

hgo<-numeric(n1)
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo

# ------------------------------------------------------------------
# obtención de las fronteras del grupo 0 del conjunto de
calibración
# ------------------------------------------------------------------
p<-numeric(n1)
for (i in 1:n1){
if (ho[i]==0 & hgo[i]==0){
p[i]<-i
}else{
}
p
}
p

g0.orig.cal<-read.table("compracal1.txt",dec=",", header=TRUE)

front0<-
rbind(g0.orig.cal[1,],g0.orig.cal[11,],g0.orig.cal[16,],g0.orig.cal
[32,],g0.orig.cal[33,],g0.orig.cal[34,],g0.orig.cal[37,])
#tener cuidado el 1 no cambia
attach(front0)
min_x1<-min(x1)
min_x3<-min(x3)
max_x7<-max(x7)
x1a<-x1-min_x1
```

```
x3a<-x3-min_x3
x7a<-max_x7-x7
f0<-rbind(x1a,x3a,x7a)
f0

#-----------------------------------------------------------------
# valores de h y hg con las fronteras originaes grupo 0, para los
dos grupos
# -----------------------------------------------------------------
setwd("D:/Desktop/datostesis")
g0.orig.cal<-read.table("compracal.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
x1a<-X1-min_x1
x3a<-X3-min_x3
x7a<-max_x7-X7
g0.cal<-rbind(x1a,x3a,x7a)

dimension<-dim(g0.cal)
n1<-dimension[2]
dimension<-dim(f0)
nf<-dimension[2]

library(boot)
h<-numeric(nf)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-g0.cal

ho<-numeric(n1)
for (i in 1:(n1)){
f0[,1]<-matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,nf)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=f0, b1=vectorTermIndep, A3=lambda,
b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
}
ho


hgo<-numeric(n1)
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo
```

```
hg1cal<-hgo
write.table(hgo, 'hg1cal.txt', sep='\t', dec=',')


#ahora lo mismo con validacion
#-----------------------------------------------------------------
setwd("D:/Desktop/datostesis")
g0.orig.cal<-read.table("compraval.txt",dec=",", header=TRUE)
attach(g0.orig.cal)
x1a<-X1-min_x1
x3a<-X3-min_x3
x7a<-max_x7-X7
g0.cal<-rbind(x1a,x3a,x7a)
dimension<-dim(g0.cal)
n1<-dimension[2]
dimension<-dim(f0)
nf<-dimension[2]

library(boot)
h<-numeric(nf)   #funcion objetivo
h[1]=1
matrizRestriccionesMenorOIgual<-g0.cal

ho<-numeric(n1)
for (i in 1:(n1)){
f0[,1]<-matrizRestriccionesMenorOIgual[,i]
vectorTermIndep <- matrizRestriccionesMenorOIgual[,i]
lambda<-matrix(1,1,nf)
lambda[1,1]<-0
lambdaTermInd<-c(1)
mySimplex <- simplex(h, A1=f0, b1=vectorTermIndep, A3=lambda,
b3=lambdaTermInd,maxi=TRUE)
ho[i]<-mySimplex$value
}
ho

hgo<-numeric(n1)
for (i in 1:n1){
hgo[i]<-((1-hmo)*ho[i])/((1-ho[i])*hmo)
}
hgo
hg1val<-hgo
write.table(hgo, 'hg1val.txt', sep='\t', dec=',')
```

11.5 Desarrollo del Modelo de Almeida paso a paso con el conjunto de datos 1 de Empresas Clientes

GRUPO O	DATOS DE CALIBRACION								
DMU	x1	x3	x7	x11	x1 ajustada	x3 ajustada	x7 ajustada	ho	hgo
Pm	2,23	6,87	8,46	0	2,27	1,43	1,86	0,4783	1,000
2	1,8	6,3	8,4	0	2,7	2	1,8	0,5844	1,534
6	1,9	7,9	9,7	0	2,6	0,4	3,1	0,3657	0,629
8	1,3	6,2	6,9	0	3,2	2,1	0,3	0,5046	1,111
13	2,8	8,1	6,6	0	1,7	0,2	0	0	0,000
24	2,4	6,7	7,2	0	2,1	1,6	0,6	0,3369	0,554
31	3	6	8	0	1,5	2,3	1,4	0,3415	0,566
36	1,8	7,5	7,6	0	2,7	0,8	1	0,4534	0,905
39	0	6,9	8,9	0	4,5	1,4	2,3	0,6756	2,272
45	2	6,5	8,5	0	2,5	1,8	1,9	0,5503	1,335
48	3,4	5,6	9,1	0	1,1	2,7	2,5	0,4393	0,855
52	2,6	8,2	9	0	1,9	0,1	2,4	0	0,000
53	4,5	6,3	8,8	0	0	2	2,2	0	0,000
54	2,8	6,7	9,2	0	1,7	1,6	2,6	0,4118	0,764
65	1,1	7,2	10	0	3,4	1,1	3,4	0,5745	1,473
68	1,6	6,4	7,1	0	2,9	1,9	0,5	0,4827	1,018
70	2,3	8,3	9,1	0	2,2	0	2,5	0	0,000
71	3,6	5,9	8,4	0	0,9	2,4	1,8	0,2579	0,379
79	1	7,1	9,9	0	3,5	1,2	3,3	0,5923	1,585
86	2,5	7	9	0	2	1,3	2,4	0,4148	0,773
89	2,9	7,3	8	0	1,6	1	1,4	0,2576	0,378
96	0,6	6,4	8,4	0	3,9	1,9	1,8	0,6683	2,198
99	3,1	6,7	8,4	0	1,4	1,6	1,8	0,3511	0,590
	MAX	MAX	MIN						
	4,5	8,3	6,6						

X1 X3 DESFAVORABLES AL GRUPO O	
X7 FAVORABLE AL GRUPO O	

GRUPO 1	DATOS DE CALIBRACION								
DMU	x1	x3	x7	x11	x1 ajustada	x3 ajustada	x7 ajustada	h1	hg1
Pm	4,26	8,57	6,01	1	2,16	3,47	2,39	0,471	1,000
1	4,1	6,9	5,2	1	2	1,8	3,2	0,447	0,905
5	6	9,6	4,5	1	3,9	4,5	3,9	0,642	2,014
7	4,6	9,5	7,6	1	2,5	4,4	0,8	0,307	0,497
11	2,4	8,8	5,8	1	0,3	3,7	2,6	0,468	0,988
12	3,9	9,1	8,3	1	1,8	4	0,1	0,041	0,048
14	3,7	8,6	6,7	1	1,6	3,5	1,7	0,373	0,666
15	4,7	9,9	6,8	1	2,6	4,8	1,6	0,470	0,993
17	3,2	5,7	6,2	1	1,1	0,6	2,2	0,057	0,067
20	4,7	9,9	6,8	1	2,6	4,8	1,6	0,470	0,993
23	3	9,1	8,4	1	0,9	4	0	0,000	0,000
25	5,1	8,7	3,8	1	3	3,6	4,6	0,649	2,069
26	4,6	7,9	4,7	1	2,5	2,8	3,7	0,560	1,425
28	5,2	9,7	6,7	1	3,1	4,6	1,7	0,468	0,985
29	3,5	9,9	5,4	1	1,4	4,8	3	0,584	1,576
32	2,8	8,9	8,2	1	0,7	3,8	0,2	0,000	0,000
33	5,2	9,3	4,6	1	3,1	4,2	3,8	0,626	1,880
42	5,9	9,6	4,5	1	3,8	4,5	3,9	0,642	2,014
43	4,9	9,3	6,2	1	2,8	4,2	2,2	0,498	1,113
47	3,1	10	3,8	1	1	4,9	4,6	0,663	2,210
49	5,8	8,8	6,7	1	3,7	3,7	1,7	0,395	0,733
50	5,4	8	5,2	1	3,3	2,9	3,2	0,523	1,229
51	3,7	8,2	5,2	1	1,6	3,1	3,2	0,522	1,226
58	5,4	9,6	7,7	1	3,3	4,5	0,7	0,298	0,476
59	4,3	7,6	4,4	1	2,2	2,5	4	0,565	1,459
61	3,1	9,9	3,8	1	1	4,8	4,6	0,661	2,182
67	4,2	9,2	7,3	1	2,1	4,1	1,1	0,332	0,557
72	5,6	8,2	5,3	1	3,5	3,1	3,1	0,525	1,237
73	3,6	9,9	4,9	1	1,5	4,8	3,5	0,617	1,804
80	4,5	8,7	6,8	1	2,4	3,6	1,6	0,369	0,657
81	5,5	8,7	4,9	1	3,4	3,6	3,5	0,583	1,570
82	3,4	5,5	6,3	1	1,3	0,4	2,1	0,000	0,000
84	2,3	7,6	7,4	1	0,2	2,5	1	0,000	0,000
88	2,1	7,4	7,2	1	0	2,3	1,2	0,000	0,000
90	4,3	9,3	7,4	1	2,2	4,2	1	0,324	0,537
92	4,8	7,6	5,8	1	2,7	2,5	2,6	0,425	0,829
93	3,1	5,1	5,9	1	1	0	2,5	0,000	0,000
95	4	6,7	5	1	1,9	1,6	3,4	0,454	0,932
97	6,1	9,2	7,1	1	4	4,1	1,3	0,368	0,653
	MIN	MIN	MAX	SUM 1					
	2,1	5,1	8,4	38					

X1 X3 FAVORABLES AL GRUPO 1
X7 DESFAVORABLE AL GRUPO 1

	FRONTERAS OIGINALES GRUPO O						
		VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	DMU	X1	X3	X7	X1	X3	X7
FR GR O	13	2,8	8,1	6,6	1,7	0,2	0
FR GR O	52	2,6	8,2	9	1,9	0,1	2,4
FR GR O	53	4,5	6,3	8,8	0	2	2,2
FR GR O	70	2,3	8,3	9,1	2,2	0	2,5
		4,5	8,3	6,6			
		MAX	MAX	MIN			

	FRONTERAS ORIGINALES GRUPO 1						
		VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	DMU	X1	X3	X7	X1	X3	X7
FR GR 1	23	3	9,1	8,4	0,9	4	0
FR GR 1	32	2,8	8,9	8,2	0,7	3,8	0,2
FR GR 1	82	3,4	5,5	6,3	1,3	0,4	2,1
FR GR 1	84	2,3	7,6	7,4	0,2	2,5	1
FR GR 1	88	2,1	7,4	7,2	0	2,3	1,2
FR GR 1	93	3,1	5,1	5,9	1	0	2,5
		2,1	5,1	8,4			
		MIN	MIN	MAX			

	DATOS DE CALIBRACION								
	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
DMU	x1	x3	x7	x11	x1	x3	x7	h	hg
2	1,8	6,3	8,4	0	2,7	2	1,8	0,584	1,534
6	1,9	7,9	9,7	0	2,6	0,4	3,1	0,366	0,629
8	1,3	6,2	6,9	0	3,2	2,1	0,3	0,505	1,111
13	2,8	8,1	6,6	0	1,7	0,2	0	0,000	0,000
24	2,4	6,7	7,2	0	2,1	1,6	0,6	0,337	0,554
31	3	6	8	0	1,5	2,3	1,4	0,342	0,566
36	1,8	7,5	7,6	0	2,7	0,8	1	0,453	0,905
39	0	6,9	8,9	0	4,5	1,4	2,3	0,676	2,272
45	2	6,5	8,5	0	2,5	1,8	1,9	0,550	1,335
48	3,4	5,6	9,1	0	1,1	2,7	2,5	0,439	0,855
52	2,6	8,2	9	0	1,9	0,1	2,4	0,000	0,000
53	4,5	6,3	8,8	0	0	2	2,2	0,000	0,000
54	2,8	6,7	9,2	0	1,7	1,6	2,6	0,412	0,764
65	1,1	7,2	10	0	3,4	1,1	3,4	0,575	1,473
68	1,6	6,4	7,1	0	2,9	1,9	0,5	0,483	1,018
70	2,3	8,3	9,1	0	2,2	0	2,5	0,000	0,000
71	3,6	5,9	8,4	0	0,9	2,4	1,8	0,258	0,379
79	1	7,1	9,9	0	3,5	1,2	3,3	0,592	1,585
86	2,5	7	9	0	2	1,3	2,4	0,415	0,773
89	2,9	7,3	8	0	1,6	1	1,4	0,258	0,378
96	0,6	6,4	8,4	0	3,9	1,9	1,8	0,668	2,198
99	3,1	6,7	8,4	0	1,4	1,6	1,8	0,351	0,590
1	4,1	6,9	5,2	1	0,4	1,4	-1,4	<0	<0
5	6	9,6	4,5	1	-1,5	-1,3	-2,1	<0	<0
7	4,6	9,5	7,6	1	-0,1	-1,2	1	<0	<0
11	2,4	8,8	5,8	1	2,1	-0,5	-0,8	<0	<0
12	3,9	9,1	8,3	1	0,6	-0,8	1,7	<0	<0
14	3,7	8,6	6,7	1	0,8	-0,3	0,1	<0	<0
15	4,7	9,9	6,8	1	-0,2	-1,6	0,2	<0	<0
17	3,2	5,7	6,2	1	1,3	2,6	-0,4	<0	<0
20	4,7	9,9	6,8	1	-0,2	-1,6	0,2	<0	<0
23	3	9,1	8,4	1	1,5	-0,8	1,8	<0	<0
25	5,1	8,7	3,8	1	-0,6	-0,4	-2,8	<0	<0
26	4,6	7,9	4,7	1	-0,1	0,4	-1,9	<0	<0
28	5,2	9,7	6,7	1	-0,7	-1,4	0,1	<0	<0
29	3,5	9,9	5,4	1	1	-1,6	-1,2	<0	<0
32	2,8	8,9	8,2	1	1,7	-0,6	1,6	<0	<0
33	5,2	9,3	4,6	1	-0,7	-1	-2	<0	<0
42	5,9	9,6	4,5	1	-1,4	-1,3	-2,1	<0	<0
43	4,9	9,3	6,2	1	-0,4	-1	-0,4	<0	<0
47	3,1	10	3,8	1	1,4	-1,7	-2,8	<0	<0
49	5,8	8,8	6,7	1	-1,3	-0,5	0,1	<0	<0
50	5,4	8	5,2	1	-0,9	0,3	-1,4	<0	<0
51	3,7	8,2	5,2	1	0,8	0,1	-1,4	<0	<0
58	5,4	9,6	7,7	1	-0,9	-1,3	1,1	<0	<0
59	4,3	7,6	4,4	1	0,2	0,7	-2,2	<0	<0
61	3,1	9,9	3,8	1	1,4	-1,6	-2,8	<0	<0
67	4,2	9,2	7,3	1	0,3	-0,9	0,7	<0	<0
72	5,6	8,2	5,3	1	-1,1	0,1	-1,3	<0	<0
73	3,6	9,9	4,9	1	0,9	-1,6	-1,7	<0	<0
80	4,5	8,7	6,8	1	0	-0,4	0,2	<0	<0
81	5,5	8,7	4,9	1	-1	-0,4	-1,7	<0	<0
82	3,4	5,5	6,3	1	1,1	2,8	-0,3	<0	<0
84	2,3	7,6	7,4	1	2,2	0,7	0,8	0,3398	0,5613947
88	2,1	7,4	7,2	1	2,4	0,9	0,6	0,4063	0,7464492
90	4,3	9,3	7,4	1	0,2	-1	0,8	<0	<0
92	4,8	7,6	5,8	1	-0,3	0,7	-0,8	<0	<0
93	3,1	5,1	5,9	1	1,4	3,2	-0,7	<0	<0
95	4	6,7	5	1	0,5	1,6	-1,6	<0	<0
97	6,1	9,2	7,1	1	-1,6	-0,9	0,5	<0	<0
Pm	2,23	6,87	8,46		2,27	1,43	1,86	0,48	1,00
								hm	hg

	DATOS DE CALIBRACION								
	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
DMU	x1	x3	x7	x11	x1	x3	x7	h	hg
2	1,8	6,3	8,4	0	-0,3	1,2	0	<0	<0
6	1,9	7,9	9,7	0	-0,2	2,8	-1,3	<0	<0
8	1,3	6,2	6,9	0	-0,8	1,1	1,5	<0	<0
13	2,8	8,1	6,6	0	0,7	3	1,8	0,3168	0,520
24	2,4	6,7	7,2	0	0,3	1,6	1,2	<0	<0
31	3	6	8	0	0,9	0,9	0,4	<0	<0
36	1,8	7,5	7,6	0	-0,3	2,4	0,8	<0	<0
39	0	6,9	8,9	0	-2,1	1,8	-0,5	<0	<0
45	2	6,5	8,5	0	-0,1	1,4	-0,1	<0	<0
48	3,4	5,6	9,1	0	1,3	0,5	-0,7	<0	<0
52	2,6	8,2	9	0	0,5	3,1	-0,6	<0	<0
53	4,5	6,3	8,8	0	2,4	1,2	-0,4	<0	<0
54	2,8	6,7	9,2	0	0,7	1,6	-0,8	<0	<0
65	1,1	7,2	10	0	-1	2,1	-1,6	<0	<0
68	1,6	6,4	7,1	0	-0,5	1,3	1,3	<0	<0
70	2,3	8,3	9,1	0	0,2	3,2	-0,7	<0	<0
71	3,6	5,9	8,4	0	1,5	0,8	0	<0	<0
79	1	7,1	9,9	0	-1,1	2	-1,5	<0	<0
86	2,5	7	9	0	0,4	1,9	-0,6	<0	<0
89	2,9	7,3	8	0	0,8	2,2	0,4	<0	<0
96	0,6	6,4	8,4	0	-1,5	1,3	0	<0	<0
99	3,1	6,7	8,4	0	1	1,6	0	<0	<0
1	4,1	6,9	5,2	1	2	1,8	3,2	0,447	0,905
5	6	9,6	4,5	1	3,9	4,5	3,9	0,642	2,014
7	4,6	9,5	7,6	1	2,5	4,4	0,8	0,307	0,497
11	2,4	8,8	5,8	1	0,3	3,7	2,6	0,468	0,988
12	3,9	9,1	8,3	1	1,8	4	0,1	0,041	0,048
14	3,7	8,6	6,7	1	1,6	3,5	1,7	0,373	0,666
15	4,7	9,9	6,8	1	2,6	4,8	1,6	0,470	0,993
17	3,2	5,7	6,2	1	1,1	0,6	2,2	0,057	0,067
20	4,7	9,9	6,8	1	2,6	4,8	1,6	0,470	0,993
23	3	9,1	8,4	1	0,9	4	0	0,000	0,000
25	5,1	8,7	3,8	1	3	3,6	4,6	0,649	2,069
26	4,6	7,9	4,7	1	2,5	2,8	3,7	0,560	1,425
28	5,2	9,7	6,7	1	3,1	4,6	1,7	0,468	0,985
29	3,5	9,9	5,4	1	1,4	4,8	3	0,584	1,576
32	2,8	8,9	8,2	1	0,7	3,8	0,2	0,000	0,000
33	5,2	9,3	4,6	1	3,1	4,2	3,8	0,626	1,880
42	5,9	9,6	4,5	1	3,8	4,5	3,9	0,642	2,014
43	4,9	9,3	6,2	1	2,8	4,2	2,2	0,498	1,113
47	3,1	10	3,8	1	1	4,9	4,6	0,663	2,210
49	5,8	8,8	6,7	1	3,7	3,7	1,7	0,395	0,733
50	5,4	8	5,2	1	3,3	2,9	3,2	0,523	1,229
51	3,7	8,2	5,2	1	1,6	3,1	3,2	0,522	1,226
58	5,4	9,6	7,7	1	3,3	4,5	0,7	0,298	0,476
59	4,3	7,6	4,4	1	2,2	2,5	4	0,565	1,459
61	3,1	9,9	3,8	1	1	4,8	4,6	0,661	2,182
67	4,2	9,2	7,3	1	2,1	4,1	1,1	0,332	0,557
72	5,6	8,2	5,3	1	3,5	3,1	3,1	0,525	1,237
73	3,6	9,9	4,9	1	1,5	4,8	3,5	0,617	1,804
80	4,5	8,7	6,8	1	2,4	3,6	1,6	0,369	0,657
81	5,5	8,7	4,9	1	3,4	3,6	3,5	0,583	1,570
82	3,4	5,5	6,3	1	1,3	0,4	2,1	0,000	0,000
84	2,3	7,6	7,4	1	0,2	2,5	1	0,000	0,000
88	2,1	7,4	7,2	1	0	2,3	1,2	0,000	0,000
90	4,3	9,3	7,4	1	2,2	4,2	1	0,324	0,537
92	4,8	7,6	5,8	1	2,7	2,5	2,6	0,425	0,829
93	3,1	5,1	5,9	1	1	0	2,5	0,000	0,000
95	4	6,7	5	1	1,9	1,6	3,4	0,454	0,932
97	6,1	9,2	7,1	1	4	4,1	1,3	0,368	0,653
Pm	4,26	8,57	6,01	1,00	2,16	3,47	2,39	0,47	1,00
								hm	hg

	CONDIC	FRONTER	FRONTER	CRITERI	RESUL-
	REAL	GRUP O	GRUP 1	MAYOR	TADO.
DMU	x11	hg	hg	hg	
2	GO	1,534	<0	**GO**	
6	GO	0,629	<0	**GO**	
8	GO	1,111	<0	**GO**	
13	GO	0,000	0,520	**G1**	**ERROR**
24	GO	0,554	<0	**GO**	
31	GO	0,566	<0	**GO**	
36	GO	0,905	<0	**GO**	
39	GO	2,272	<0	**GO**	
45	GO	1,335	<0	**GO**	
48	GO	0,855	<0	**GO**	
52	GO	0,000	<0	**GO**	
53	GO	0,000	<0	**GO**	
54	GO	0,764	<0	**GO**	
65	GO	1,473	<0	**GO**	
68	GO	1,018	<0	**GO**	
70	GO	0,000	<0	**GO**	
71	GO	0,379	<0	**GO**	
79	GO	1,585	<0	**GO**	
86	GO	0,773	<0	**GO**	
89	GO	0,378	<0	**GO**	
96	GO	2,198	<0	**GO**	
99	GO	0,590	<0	**GO**	
1	G1	<0	0,905	**G1**	
5	G1	<0	2,014	**G1**	
7	G1	<0	0,497	**G1**	
11	G1	<0	0,988	**G1**	
12	G1	<0	0,048	**G1**	
14	G1	<0	0,666	**G1**	
15	G1	<0	0,993	**G1**	
17	G1	<0	0,067	**G1**	
20	G1	<0	0,993	**G1**	
23	G1	<0	0,000	**G1**	
25	G1	<0	2,069	**G1**	
26	G1	<0	1,425	**G1**	
28	G1	<0	0,985	**G1**	
29	G1	<0	1,576	**G1**	
32	G1	<0	0,000	**G1**	
33	G1	<0	1,880	**G1**	
42	G1	<0	2,014	**G1**	
43	G1	<0	1,113	**G1**	
47	G1	<0	2,210	**G1**	
49	G1	<0	0,733	**G1**	
50	G1	<0	1,229	**G1**	
51	G1	<0	1,226	**G1**	
58	G1	<0	0,476	**G1**	
59	G1	<0	1,459	**G1**	
61	G1	<0	2,182	**G1**	
67	G1	<0	0,557	**G1**	
72	G1	<0	1,237	**G1**	
73	G1	<0	1,804	**G1**	
80	G1	<0	0,657	**G1**	
81	G1	<0	1,570	**G1**	
82	G1	<0	0,000	**G1**	
84	G1	0,561	0,000	**GO**	**ERROR**
88	G1	0,746	0,000	**GO**	**ERROR**
90	G1	<0	0,537	**G1**	
92	G1	<0	0,829	**G1**	
93	G1	<0	0,000	**G1**	
95	G1	<0	0,932	**G1**	
97	G1	<0	0,653	**G1**	

	FRONTERAS ORIGINALES GRUPO O						
		VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	DMU	X1	X3	X7	X1	X3	X7
FR GR O	13	2,8	8,1	6,6	1,7	0,2	0
FR GR O	52	2,6	8,2	9	1,9	0,1	2,4
FR GR O	53	4,5	6,3	8,8	0	2	2,2
FR GR O	70	2,3	8,3	9,1	2,2	0	2,5
		4,5	8,3	6,6			
		MAX	MAX	MIN			

	FRONTERAS ORIGINALES GRUPO 1						
		VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	DMU	X1	X3	X7	X1	X3	X7
FR GR 1	23	3	9,1	8,4	0,9	4	0
FR GR 1	32	2,8	8,9	8,2	0,7	3,8	0,2
FR GR 1	82	3,4	5,5	6,3	1,3	0,4	2,1
FR GR 1	84	2,3	7,6	7,4	0,2	2,5	1
FR GR 1	88	2,1	7,4	7,2	0	2,3	1,2
FR GR 1	93	3,1	5,1	5,9	1	0	2,5
		2,1	5,1	8,4			
		MIN	MIN	MAX			

	DATOS DE VALIDACION								
DMU	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
	x1	x3	x7	x11	x1	x3	x7	h	hg
3	3,4	5,7	8,2	0	1,1	2,6	1,6	0,272	0,408
4	2,7	7,1	7,8	0	1,8	1,2	1,2	0,356	0,603
10	4	6,5	8,7	0	0,5	1,8	2,1	0,141	0,180
27	2,4	6,6	7,2	0	2,1	1,7	0,6	0,337	0,554
30	4,1	5,9	8,4	0	0,4	2,4	1,8	0,051	0,058
34	3,4	6,4	8,4	0	1,1	1,9	1,8	0,318	0,507
35	2,4	7,7	6,2	0	2,1	0,6	-0,4	<0	<0
37	3,6	5,8	9,3	0	0,9	2,5	2,7	0,421	0,792
40	2,4	6,4	8,8	0	2,1	1,9	2,2	0,515	1,158
41	1,9	7,6	7,7	0	2,6	0,7	1,1	0,421	0,792
57	4,9	7,4	9,6	0	-0,4	0,9	3	<0	<0
60	2,3	8	8,7	0	2,2	0,3	2,1	0,239	0,343
75	3	6,6	8,2	0	1,5	1,7	1,6	0,379	0,665
83	1,6	6,1	8,2	0	2,9	2,2	1,6	0,589	1,563
85	2,6	8,5	6,8	0	1,9	-0,2	0,2	<0	<0
87	2,4	8,4	6,7	0	2,1	-0,1	0,1	<0	<0
94	1,9	5	8,2	0	2,6	3,3	1,6	0,557	1,371
98	2	5,2	8,4	0	2,5	3,1	1,8	0,563	1,406
9	5,5	9,4	7,6	1	-1	-1,1	1	<0	<0
16	3,4	9,7	4,8	1	1,1	-1,4	-1,8	<0	<0
18	4,9	7,7	5,9	1	-0,4	0,6	-0,7	<0	<0
19	5,3	9,7	6,8	1	-0,8	-1,4	0,2	<0	<0
21	3,3	8,6	6,3	1	1,2	-0,3	-0,3	<0	<0
22	3,4	8,3	5,2	1	1,1	0	-1,4	<0	<0
38	4	9,1	7,3	1	0,5	-0,8	0,7	<0	<0
44	5	8,6	3,7	1	-0,5	-0,3	-2,9	<0	<0
46	5	9,4	6,3	1	-0,5	-1,1	-0,3	<0	<0
55	3,8	8,7	5,6	1	0,7	-0,4	-1	<0	<0
56	2,9	7,7	7,7	1	1,6	0,6	1,1	0,128	0,160
62	5,1	9,2	4,5	1	-0,6	-0,9	-2,1	<0	<0
63	4,1	9,3	7,4	1	0,4	-1	0,8	<0	<0
64	3	5,5	6	1	1,5	2,8	-0,6	<0	<0
66	3,7	9	6,8	1	0,8	-0,7	0,2	<0	<0
69	5,3	8,5	4,8	1	-0,8	-0,2	-1,8	<0	<0
74	5,2	9,1	7,3	1	-0,7	-0,8	0,7	<0	<0
76	4,2	9,4	8,5	1	0,3	-1,1	1,9	<0	<0
77	3,8	8,3	5,3	1	0,7	0	-1,3	<0	<0
78	3,3	9,7	5,2	1	1,2	-1,4	-1,4	<0	<0
91	3	7,8	7,9	1	1,5	0,5	1,3	0,042	0,048
100	2,5	9	6	1	2	-0,7	-0,6	<0	<0
Pm	2,23	6,87	8,46		2,27	1,43	1,86	0,48	1
								hm	hg

	DATOS DE VALIDACION								
	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
DMU	x1	x3	x7	x11	x1	x3	x7	h	hg
3	3,4	5,7	8,2	0	1,3	0,6	0,2	-3,2424	-0,857
4	2,7	7,1	7,8	0	0,6	2	0,6	-0,3405	-0,285
10	4	6,5	8,7	0	1,9	1,4	-0,3	<0	<0
27	2,4	6,6	7,2	0	0,3	1,5	1,2	-0,1962	-0,184
30	4,1	5,9	8,4	0	2	0,8	0	-4	-0,897
34	3,4	6,4	8,4	0	1,3	1,3	0	-2,0769	-0,757
35	2,4	7,7	6,2	0	0,3	2,6	2,2	0,3009	0,483
37	3,6	5,8	9,3	0	1,5	0,7	-0,9	<0	<0
40	2,4	6,4	8,8	0	0,3	1,3	-0,4	<0	<0
41	1,9	7,6	7,7	0	-0,2	2,5	0,7	<0	<0
57	4,9	7,4	9,6	0	2,8	2,3	-1,2	<0	<0
60	2,3	8	8,7	0	0,2	2,9	-0,3	<0	<0
75	3	6,6	8,2	0	0,9	1,5	0,2	-1,1705	-0,605
83	1,6	6,1	8,2	0	-0,5	1	0,2	<0	<0
85	2,6	8,5	6,8	0	0,5	3,4	1,6	0,3188	0,525
87	2,4	8,4	6,7	0	0,3	3,3	1,7	0,3154	0,517
94	1,9	5	8,2	0	-0,2	-0,1	0,2	<0	<0
98	2	5,2	8,4	0	-0,1	0,1	0	<0	<0
9	5,5	9,4	7,6	1	3,4	4,3	0,8	0,2947	0,469
16	3,4	9,7	4,8	1	1,3	4,6	3,6	0,612	1,769
18	4,9	7,7	5,9	1	2,8	2,6	2,5	0,4191	0,809
19	5,3	9,7	6,8	1	3,2	4,6	1,6	0,4553	0,937
21	3,3	8,6	6,3	1	1,2	3,5	2,1	0,4218	0,818
22	3,4	8,3	5,2	1	1,3	3,2	3,2	0,521	1,220
38	4	9,1	7,3	1	1,9	4	1,1	0,3204	0,529
44	5	8,6	3,7	1	2,9	3,5	4,7	0,6498	2,081
46	5	9,4	6,3	1	2,9	4,3	2,1	0,4937	1,093
55	3,8	8,7	5,6	1	1,7	3,6	2,8	0,5143	1,187
56	2,9	7,7	7,7	1	0,8	2,6	0,7	-0,0647	-0,068
62	5,1	9,2	4,5	1	3	4,1	3,9	0,6278	1,891
63	4,1	9,3	7,4	1	2	4,2	1	0,3237	0,537
64	3	5,5	6	1	0,9	0,4	2,4	0,0507	0,060
66	3,7	9	6,8	1	1,6	3,9	1,6	0,3937	0,728
69	5,3	8,5	4,8	1	3,2	3,4	3,6	0,5821	1,562
74	5,2	9,1	7,3	1	3,1	4	1,1	0,3204	0,529
76	4,2	9,4	8,5	1	2,1	4,3	-0,1	<0	<0
77	3,8	8,3	5,3	1	1,7	3,2	3,1	0,5198	1,214
78	3,3	9,7	5,2	1	1,2	4,6	3,2	0,5878	1,599
91	3	7,8	7,9	1	0,9	2,7	0,5	-0,1245	-0,124
100	2,5	9	6	1	0,4	3,9	2,4	0,4658	0,978
Pm	4,26	8,57	6,01	1,00	2,16	3,47	2,39	0,47	1,000
								hm	hg

	CONDIC	FRONTER	FRONTER	CRITERI	RESUL-
	REAL	GRUP O	GRUP 1	MAYOR	TADO.
DMU	x11	hg	hg	hg	
3	GO	0,408	-0,857	GO	
4	GO	0,603	-0,285	GO	
10	GO	0,180	<0	GO	
27	GO	0,554	-0,184	GO	
30	GO	0,058	-0,897	GO	
34	GO	0,507	-0,757	GO	
35	GO	<0	0,483	G1	ERROR
37	GO	0,792	<0	GO	
40	GO	1,158	<0	GO	
41	GO	0,792	<0	GO	
57	GO	<0	<0	*	SIN CLASIF
60	GO	0,343	<0	GO	
75	GO	0,665	-0,605	GO	
83	GO	1,563	<0	GO	
85	GO	<0	0,525	G1	ERROR
87	GO	<0	0,517	G1	ERROR
94	GO	1,371	<0	GO	
98	GO	1,406	<0	GO	
9	G1	<0	0,469	G1	
16	G1	<0	1,769	G1	
18	G1	<0	0,809	G1	
19	G1	<0	0,937	G1	
21	G1	<0	0,818	G1	
22	G1	<0	1,220	G1	
38	G1	<0	0,529	G1	
44	G1	<0	2,081	G1	
46	G1	<0	1,093	G1	
55	G1	<0	1,187	G1	
56	G1	0,160	-0,068	G0	ERROR
62	G1	<0	1,891	G1	
63	G1	<0	0,537	G1	
64	G1	<0	0,060	G1	
66	G1	<0	0,728	G1	
69	G1	<0	1,562	G1	
74	G1	<0	0,529	G1	
76	G1	<0	<0	*	SIN CLASIF
77	G1	<0	1,214	GO	
78	G1	<0	1,599	GO	
91	G1	0,048	-0,124	G1	ERROR
100	G1	<0	0,978	GO	

	FRONTERA ACTUALIZADA GRUPO 0							
			VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	MUESTRA	DMU	X1	X3	X7	X1	X3	X7
FR GO ACT	CALIBRAC	13	2,8	8,1	6,6	2,1	0,2	0,0
FR GO ACT	CALIBRAC	52	2,6	8,2	9	2,3	0,1	2,4
FR GO ACT	CALIBRAC	53	4,5	6,3	8,8	0,4	2,0	2,2
FR GO ACT	CALIBRAC	70	2,3	8,3	9,1	2,6	0,0	2,5
FR GO ACT	VALIDAC	57	4,9	7,4	9,6	0,0	0,9	3,0
			4,90	8,30	6,60			
			MAX	MAX	MIN			

	FRONTERA ACTUALIZADA GRUPO 1							
			VARIABLES ORIGINALES			VARIABLES AJUSTADAS		
	MUESTRA	DMU	X1	X3	X7	X1	X3	X7
FR GR1 ACT	CALIBRAC	23	3	9,1	8,4	0,9	4	0,1
FR GR1 ACT	CALIBRAC	32	2,8	8,9	8,2	0,7	3,8	0,3
FR GR1 ACT	CALIBRAC	82	3,4	5,5	6,3	1,3	0,4	2,2
FR GR1 ACT	CALIBRAC	84	2,3	7,6	7,4	0,2	2,5	1,1
FR GR1 ACT	CALIBRAC	88	2,1	7,4	7,2	0	2,3	1,3
FR GR1 ACT	CALIBRAC	93	3,1	5,1	5,9	1	0	2,6
FR GR1 ACT	VALIDAC	76	4,2	9,4	8,5	2,1	4,3	0
			2,1	5,1	8,5			
			MIN	MIN	MAX			

	DATOS DE VALIDACION								
DMU	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
	x1	x3	x7	x11	x1	x3	x7	h	hg
3	3,4	5,7	8,2	0	1,5	2,6	1,6	0,232	0,329
4	2,7	7,1	7,8	0	2,2	1,2	1,2	0,326	0,529
10	4	6,5	8,7	0	0,9	1,8	2,1	0,161	0,209
27	2,4	6,6	7,2	0	2,5	1,7	0,6	0,291	0,449
30	4,1	5,9	8,4	0	0,8	2,4	1,8	0,042	0,047
34	3,4	6,4	8,4	0	1,5	1,9	1,8	0,274	0,411
35	2,4	7,7	6,2	0	2,5	0,6	-0,4	<0	<0
37	3,6	5,8	9,3	0	1,3	2,5	2,7	0,379	0,666
40	2,4	6,4	8,8	0	2,5	1,9	2,2	0,496	1,072
41	1,9	7,6	7,7	0	3,0	0,7	1,1	0,445	0,873
57	4,9	7,4	9,6	0	0,0	0,9	3,0	0,000	0,000
60	2,3	8	8,7	0	2,6	0,3	2,1	0,244	0,351
75	3	6,6	8,2	0	1,9	1,7	1,6	0,330	0,538
83	1,6	6,1	8,2	0	3,3	2,2	1,6	0,537	1,266
85	2,6	8,5	6,8	0	2,3	-0,2	0,2	<0	<0
87	2,4	8,4	6,7	0	2,5	-0,1	0,1	<0	<0
94	1,9	5	8,2	0	3,0	3,3	1,6	0,504	1,110
98	2	5,2	8,4	0	2,9	3,1	1,8	0,511	1,138
9	5,5	9,4	7,6	1	-0,6	-1,1	1,0	<0	<0
16	3,4	9,7	4,8	1	1,5	-1,4	-1,8	<0	<0
18	4,9	7,7	5,9	1	0,0	0,6	-0,7	<0	<0
19	5,3	9,7	6,8	1	-0,4	-1,4	0,2	<0	<0
21	3,3	8,6	6,3	1	1,6	-0,3	-0,3	<0	<0
22	3,4	8,3	5,2	1	1,5	0,0	-1,4	<0	<0
38	4	9,1	7,3	1	0,9	-0,8	0,7	<0	<0
44	5	8,6	3,7	1	-0,1	-0,3	-2,9	<0	<0
46	5	9,4	6,3	1	-0,1	-1,1	-0,3	<0	<0
55	3,8	8,7	5,6	1	1,1	-0,4	-1,0	<0	<0
56	2,9	7,7	7,7	1	2,0	0,6	1,1	0,245	0,353
62	5,1	9,2	4,5	1	-0,2	-0,9	-2,1	<0	<0
63	4,1	9,3	7,4	1	0,8	-1,0	0,8	<0	<0
64	3	5,5	6	1	1,9	2,8	-0,6	<0	<0
66	3,7	9	6,8	1	1,2	-0,7	0,2	<0	<0
69	5,3	8,5	4,8	1	-0,4	-0,2	-1,8	<0	<0
74	5,2	9,1	7,3	1	-0,3	-0,8	0,7	<0	<0
76	4,2	9,4	8,5	1	0,7	-1,1	1,9	<0	<0
77	3,8	8,3	5,3	1	1,1	0,0	-1,3	<0	<0
78	3,3	9,7	5,2	1	1,6	-1,4	-1,4	<0	<0
91	3	7,8	7,9	1	1,9	0,5	1,3	0,208	0,286
100	2,5	9	6	1	2,4	-0,7	-0,6	<0	<0
Pm	2,23	6,87	8,46		2,27	1,43	1,86	0,48	1,000
								hm	hg

	DATOS DE VALIDACION								
DMU	VARIABLES ORIGINALES				VARIABLES AJUSTADAS				
	x1	x3	x7	x11	x1	x3	x7	h	hg
3	3,4	5,7	8,2	0	1,3	0,6	0,3	<0	<0
4	2,7	7,1	7,8	0	0,6	2	0,7	<0	<0
10	4	6,5	8,7	0	1,9	1,4	-0,2	<0	<0
27	2,4	6,6	7,2	0	0,3	1,5	1,3	<0	<0
30	4,1	5,9	8,4	0	2	0,8	0,1	<0	<0
34	3,4	6,4	8,4	0	1,3	1,3	0,1	<0	<0
35	2,4	7,7	6,2	0	0,3	2,6	2,3	0,301	0,469
37	3,6	5,8	9,3	0	1,5	0,7	-0,8	<0	<0
40	2,4	6,4	8,8	0	0,3	1,3	-0,3	<0	<0
41	1,9	7,6	7,7	0	-0,2	2,5	0,8	<0	<0
57	4,9	7,4	9,6	0	2,8	2,3	-1,1	<0	<0
60	2,3	8	8,7	0	0,2	2,9	-0,2	<0	<0
75	3	6,6	8,2	0	0,9	1,5	0,3	<0	<0
83	1,6	6,1	8,2	0	-0,5	1	0,3	<0	<0
85	2,6	8,5	6,8	0	0,5	3,4	1,7	0,310	0,491
87	2,4	8,4	6,7	0	0,3	3,3	1,8	0,308	0,485
94	1,9	5	8,2	0	-0,2	-0,1	0,3	<0	<0
98	2	5,2	8,4	0	-0,1	0,1	0,1	<0	<0
9	5,5	9,4	7,6	1	3,4	4,3	0,9	0,286	0,437
16	3,4	9,7	4,8	1	1,3	4,6	3,7	0,604	1,661
18	4,9	7,7	5,9	1	2,8	2,6	2,6	0,409	0,755
19	5,3	9,7	6,8	1	3,2	4,6	1,7	0,445	0,874
21	3,3	8,6	6,3	1	1,2	3,5	2,2	0,412	0,766
22	3,4	8,3	5,2	1	1,3	3,2	3,3	0,512	1,145
38	4	9,1	7,3	1	1,9	4	1,2	0,311	0,493
44	5	8,6	3,7	1	2,9	3,5	4,8	0,641	1,947
46	5	9,4	6,3	1	2,9	4,3	2,2	0,483	1,020
55	3,8	8,7	5,6	1	1,7	3,6	2,9	0,505	1,111
56	2,9	7,7	7,7	1	0,8	2,6	0,8	<0	<0
62	5,1	9,2	4,5	1	3	4,1	4	0,619	1,770
63	4,1	9,3	7,4	1	2	4,2	1,1	0,315	0,501
64	3	5,5	6	1	0,9	0,4	2,5	0,049	0,056
66	3,7	9	6,8	1	1,6	3,9	1,7	0,385	0,681
69	5,3	8,5	4,8	1	3,2	3,4	3,7	0,572	1,457
74	5,2	9,1	7,3	1	3,1	4	1,2	0,311	0,493
76	4,2	9,4	8,5	1	2,1	4,3	0	0,000	0,000
77	3,8	8,3	5,3	1	1,7	3,2	3,2	0,510	1,136
78	3,3	9,7	5,2	1	1,2	4,6	3,3	0,578	1,496
91	3	7,8	7,9	1	0,9	2,7	0,6	<0	<0
100	2,5	9	6	1	0,4	3,9	2,5	0,457	0,918
Pm	2,23	6,87	8,46		2,27	1,43	1,86	0,48	1,000
								hm	hg

	FRONTERA ACTUALIZADA MUESTRA VALIDACION				
	CONDIC	FRONTER	FRONTER	CRITERI	RESUL-
	REAL	GRUP O	GRUP 1	MAYOR	TADO.
DMU	x11	hg	hg	hg	
3	GO	0,33	<0	GO	
4	GO	0,53	<0	GO	
10	GO	0,21	<0	GO	
27	GO	0,45	<0	GO	
30	GO	0,05	<0	GO	
34	GO	0,41	<0	GO	
35	GO	<0	0,47	G1	ERROR
37	GO	0,67	<0	GO	
40	GO	1,07	<0	GO	
41	GO	0,87	<0	GO	
57	GO	0,00	<0	GO	
60	GO	0,35	<0	GO	
75	GO	0,54	<0	GO	
83	GO	1,27	<0	GO	
85	GO	<0	0,49	G1	ERROR
87	GO	<0	0,49	G1	ERROR
94	GO	1,11	<0	GO	
98	GO	1,14	<0	GO	
9	G1	<0	0,44	G1	
16	G1	<0	1,66	G1	
18	G1	<0	0,75	G1	
19	G1	<0	0,87	G1	
21	G1	<0	0,77	G1	
22	G1	<0	1,15	G1	
38	G1	<0	0,49	G1	
44	G1	<0	1,95	G1	
46	G1	<0	1,02	G1	
55	G1	<0	1,11	G1	
56	G1	0,35	<0	GO	ERROR
62	G1	<0	1,77	G1	
63	G1	<0	0,50	G1	
64	G1	<0	0,06	G1	
66	G1	<0	0,68	G1	
69	G1	<0	1,46	G1	
74	G1	<0	0,49	G1	
76	G1	<0	0,00	G1	
77	G1	<0	1,14	G1	
78	G1	<0	1,50	G1	
91	G1	0,29	<0	GO	ERROR
100	G1	<0	0,92	G1	